



The Next Wave

The National Security Agency's Review of Emerging Technologies

Vol 17 No 2 • 2008

processing

verify

An Information Visualization Primer and Field Trip

Visual Analytics – Adventures in Exploring Information Spaces

Do You Know What's Happening – Just Beyond Your View?

Knowledge Visualization with Concept Mapping Tool



Letter from the Editor

*A picture is worth a thousand words.
We need to see the big picture.*

As these clichés indicate, pictures help us understand concepts and ideas that words cannot convey—no matter how many words we use. Pictures help us make mental leaps and enable us to *see things differently*. Combining this human need to use visuals as short-cuts for ideas and the deluge of data we face at work and at home, the field of data visualization helps us make sense of our fast-paced world.

In our second issue of 2008, *The Next Wave* (TNW) presents articles on several aspects of the field of data visualization. Our five articles include:

- an information visualization primer and “tour” of the Internet
- an explanation of visual analytics, an emerging field that focuses on injecting interactive visualization throughout the analysis process
- a description of longshot displays, a specific visualization technique that is used to maintain information awareness or context in a display showing large amounts of data
- an introduction to concept mapping as a technique and as a set of computer-based tools for capturing, representing, and sharing knowledge
- a brief look at visualization in virtual worlds, specifically in Second Life

We hope you enjoy the issue and come away with a *big picture* of the field of data visualization.

The Next Wave is published to disseminate significant technical advancements and research activities in telecommunications and information technologies. Mentions of company names or commercial products do not imply endorsement by the U.S. Government. Articles present views of the authors and not necessarily those of NSA or the TNW staff.



For more information, please contact us at
TNW@tycho.nsc.mil





CONTENTS

FEATURES

- 4 An Information Visualization Primer and Field Trip
- 14 Visual Analytics: Adventures in Exploring Information
- 20 Do You Know What's Happening – Just Beyond Your View?
- 26 Knowledge Visualization with Concept Mapping Tool

FOCUS

- 30 Visualizing Data in the Metaverse
-



An Information Visualization Primer and Field Trip

The commonplace concepts of information explosion and information overload... the calls to convert data into usable information and knowledge... the rallying cry “Volume, velocity, variety!” that describes an ever-increasing amount of data—all of these refer to the challenge of analyzing and understanding data. Terms, for example, *information economy* and *knowledge economy*, have been coined to refer to the world’s post-industrial economies, racing to understand and profit from their data.

On the large scale, our nation faces the globalizing influence of the Internet and the access to more data, which is fueling a drive to collect and capitalize on the data. On a smaller scale, you may face the task of analyzing several months of data and writing a report on your conclusions. To do so, you must query several different databases, normalize and collect the data together, identify trends and outliers, produce understandable, synthesized conclusions... all in time, you hope, to get home for the season premiere of *Lost*.

What Is Visualization?

Visualization is one strategy for making sense of data. For the remainder of this article, we will consider whether visualization could help you work more efficiently and effectively. The article focuses on the goal of understanding data, especially when the understanding relates to a task that must be performed. This article does not treat the topic of visualization used for art or entertainment, but rather for analytic advantage and scientific breakthrough.

What constitutes visualization? Some authors, defining visualization, focus on internal human processes. They discuss

visualization as the human, mental activity starting from perception and leading to a mental model in memory that represents some level of understanding. Moving outside of the human, this article examines how technology can serve as external aids in the formation of human comprehension. From the viewpoint that technology can help us to be smart, this article provides this commonly used definition that neatly captures both the purpose and the method of visualization:

Visualization—The use of computer-supported, interactive, visual representations of data to amplify cognition.” [1]

A related definition is offered by the National Visualization and Analytics Center (NVAC):

Visual Analytics—“The science of analytical reasoning facilitated by interactive visual interfaces.” [2]

Visualization is a computerized interface using visual cues to improve human understanding of data. Visual cues can range from points, lines, shapes, colors, intensity, textures, motion—anything that can be used to encode aspects of the data to make the data more accessible to human perception. Effective visualization means selecting and organizing the visual cues so that you can do your work faster or with fewer mistakes or possibly with

discoveries that you would not otherwise make.

Within the field of visualization, a distinction has historically been made between **scientific visualization** and **information visualization**. The distinction revolves around the type of data to be visualized. Scientific visualization is based on physical data about, for example, molecules or the weather. This type of data generally has a natural visual representation because the data reflect an existing three-dimensional reality. Examples of scientific visualization include climate studies, simulations of weapons testing, models of chemical or biological systems, air flow dynamics of vehicles, and structural engineering models. In contrast, information visualization is based on abstract data that, consequently, do not have an obvious visual representation. Examples of information visualization include stock market views, sports statistics analysis, human genome mapping, the architecture of a software system, and social network graphs. This article focuses on information visualization.

Repeating from above, visualization is one strategy for making sense of data. However, there are other strategies, and visualization is almost always used in conjunction with strategies from information retrieval, data correlation/fusion, interaction techniques, and analytic algorithms (e.g., data mining approaches, clustering techniques, and natural language processing). Visualization, per se, is not a complete solution. A complete solution includes a way to collect and store data, fast query mechanisms, possibly some automated analysis, and an interface that gives the human maximum flexibility in examining the data and automated analysis.

A common combination is to mix visualization with interaction techniques to produce **interactive visualization**. Interaction techniques include motions such as mouse clicks, mouse dragging, or finger actions on touch screens. The use of direct interactions to control a visualization results in dynamic visualizations that change as a user directly

State	Average Residential Electricity Cost (cents per kilowatt hour)	Adopted Electricity Restructuring	State	Average Residential Electricity Cost (cents per kilowatt hour)	Adopted Electricity Restructuring
Alabama	8.23	no	Montana	8.22	yes
Alaska	13.71	no	Nebraska	6.55	no
Arizona	8.17	yes	Nevada	10.88	no
Arkansas	7.78	no	Nevada	15.02	yes
California	13.47	no	New Hampshire	11.38	yes
Colorado	9.2	no	New Jersey	9.01	no
Connecticut	16.26	yes	New Mexico	16.52	yes
Delaware	8.66	yes	New York	8.95	no
Florida	11.14	no	North Carolina	6.45	no
Georgia	8.71	no	North Dakota	8.94	yes
Hawaii	22.94	no	Ohio	8.46	no
Idaho	6.14	no	Oklahoma	7.46	no
Illinois	8.05	yes	Oregon	10.05	yes
Indiana	7.77	no	Pennsylvania	14.73	yes
Iowa	9.68	no	Rhode Island	8.82	no
Kansas	7.43	no	South Carolina	7.33	no
Kentucky	6.53	no	South Dakota	6.53	no
Louisiana	9.59	no	Tennessee	11.85	yes
Maine	14.07	yes	Texas	7.35	no
Maryland	7.93	yes	Utah	13.36	no
Massachusetts	18.43	yes	Vermont	7.94	yes
Michigan	9.56	yes	Virginia	6.67	no
Minnesota	8.63	no	Washington	6.15	no
Mississippi	9.66	no	Wisconsin	10.2	no
Missouri	6.55	no	Wyoming	7.07	no
			District of Columbia	7.89	yes

Figure 1:
Data about
1) the average
cost of electricity
for the 50 US
states and
the District of
Columbia and
2) which states
have adopted
electricity
restructuring

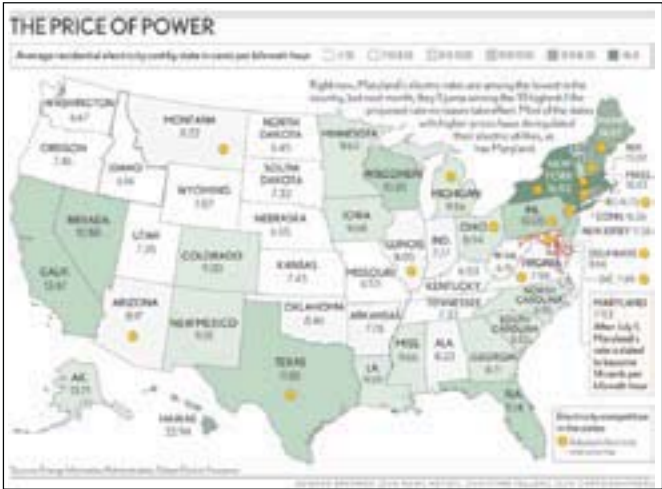


Figure 2:
Published by the
Baltimore Sun,
this map displays
the data from
Figure 1.
Average electricity
cost appears in
each state.
Yellow circles mark
the states that have
adopted electricity
restructuring.

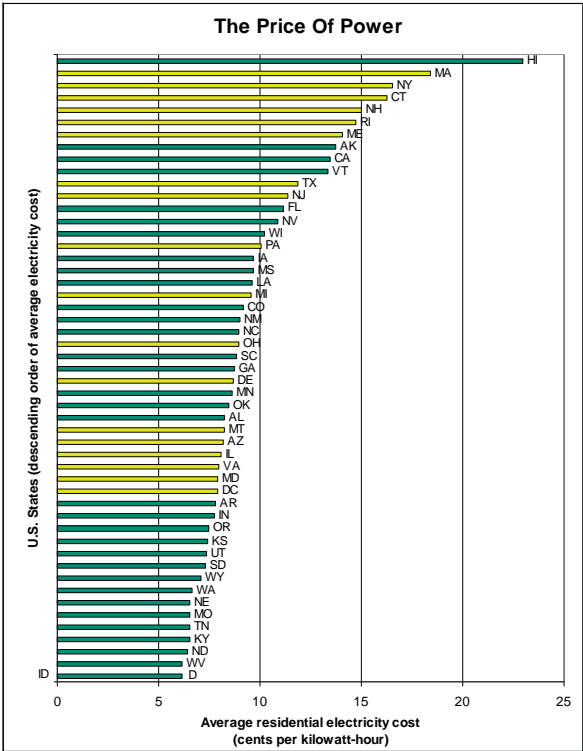


Figure 3:
Bar chart showing
average residential
electricity cost for
the 50 US states
and the District of
Columbia.
The states marked
in yellow have
adopted electricity
restructuring.

manipulates the visualization. Interactive visualization functions as both input and output mechanisms between the human and machine. As a subtle, final point in defining visualization, you can think of interactive visualization as a database interface. The interaction techniques enable “query by direct manipulation,” which is hopefully simpler than writing an equivalent database query. This combination is especially helpful because information overload results not only from increasing volumes of data, but also from the increasing complexity of the questions that we want to ask.

Why Use Visualization?

Why might one prefer a visual representation to a textual representation or an analytic algorithm? Answers like “A picture is worth a thousand words” or “The eye is the highest bandwidth to the brain” do not suffice. These vague, uninformative statements appear in marketing, for example, but provide no actual explanation for an inquisitive reader. Let’s take a more precise stab at the question using both an inductive example and then a more detailed answer.

As our example, I will describe a data set and a task that you must complete. There are three representations of the data offered to you, and you can choose one of the three to help you complete the task. The data concern the cost of electricity in the fifty US states and the District of Columbia and also which states have opted for electricity restructuring. Your task is to answer this question: Is there a relationship between electricity restructuring and the cost of electricity? Figure 1 shows the full data

set as a textual table. Using Figure 1, how would you proceed to solve your assigned task? Figure 2 shows the same data in a format that was published in the Baltimore Sun, in June 2006, as a geographical map with color coding [3]. Figure 3 shows the data set as a bar chart with color coding. Which data representation do you want to use and why... the tabular form, the colored map, or the bar chart?

Your choice of representation likely gives you an example showing how data representation can make a difference. Your decision may involve a number of observations. For instance, you may note that the table is alphabetized by state name, whereas you might want it to be ordered by a different attribute of the data. You may note that the map provides a familiar orientation but that the geographical orientation does not directly pertain to your task. In the end, it is unlikely that you will assess these three data representations to be equal. You will likely prefer one of the representations because you judge it to “work better” for your assigned task.

This specific example is, in fact, consistent with cognitive studies of how representation affects people’s comprehension and problem solving skills. Looking at the cognitive and human factors literature, you can find many examples where one representation is better or worse than another for a specific task. In designing an interface, therefore, success occurs when the designer creates an effective representation for the intended tasks.

So representation can make a difference in performance, but why would a visual representation be helpful? The crux

of this answer lies in the fact that humans process, even with brief exposure, certain visual cues very quickly. Perception is the visual process by which people initially see their world. Perception is not actually a single mental system, but rather comprises a number of vision subsystems. As a simplification, you can think of perception as a two-stage process. In the first stage, visual information is processed rapidly and in parallel to identify basic features of the external visual frame. In the second stage, visual attention, which is largely a function of eye movement, directs processing, and items in the external visual frame are processed sequentially.

The full mechanisms involved in human perception and cognition are not fully understood, but some generally accepted concepts do exist that help to explain why visualization may be helpful. In the first stage of perception, we know that certain features pop out to a human eye, including 2D spatial position, color, size, orientation, and shape. This theoretical mechanism is called pre-attentive processing, because the work occurs rapidly (on the order of 10 milliseconds per visual item) before conscious attention occurs. To illustrate this effect, consider how the task of finding the occurrences of the 2s in the numbers below is a tedious prospect:

712985791828759871983749812723
958719873498719384719983798134958
639847937558783449867873647658365
847594060374961837173587834587349
560346807740759189638738761475983
374068340576818734591874598373698
239898569833749573985791847868579
040795806597037918618726981649867

In contrast, when highlighted for pre-attentive processing, the relevant digits appear before you pay conscious attention:

71**2**9857918**2**875987198374981**2**723
958719873498719384719983798134958
639847937558783449867873647658365
847594060374961837173587834587349
560346807740759189638738761475983
374068340576818734591874598373698
239898569833749573985791847868579
0407958065970379186187**2**6981649867

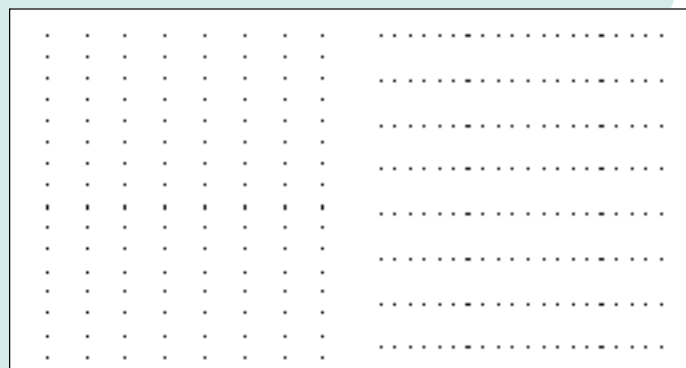


Figure 4:
An example of the principle of proximity. Because of spatial placement, people tend to perceive the dots on the left as columns and the dots on the right as rows.

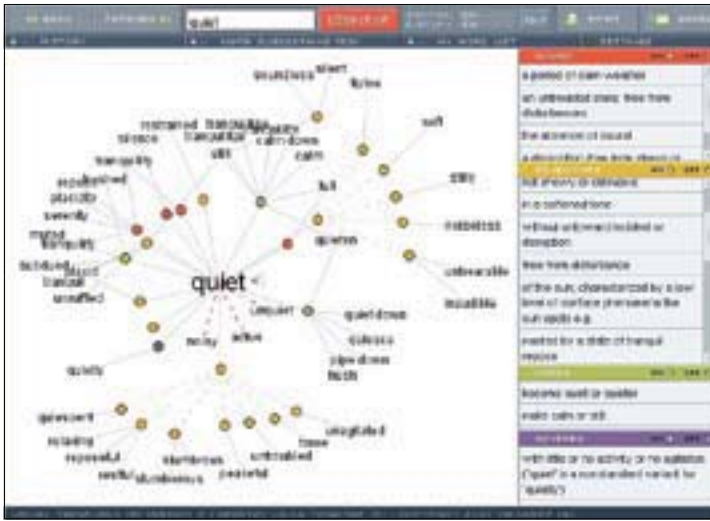


Figure 5: Visual Thesaurus display for the word *Quiet*



Figure 6: Inspecting one definition of *Quiet* (Visual Thesaurus)

Image or text from the Visual Thesaurus (<http://www.visualthesaurus.com>), copyright 1998-2008 Thinkmap, Inc. All rights reserved.

Going beyond the first stage of perception, we also know that humans engage in object perception, which includes perceiving that several visual features are related and form a group. Building from basic visual elements such as color and stereo depth, people build contours and boundaries and use these to segment their visual world into regions. The Gestalt principles, as related to perception, describe how people perceive visual elements as organized into patterns. For example, the principle of proximity captures the fact that, when people see objects that are close together, they perceive the objects as one group (see Figure 4).

The rationale for visualization is that judicious use of visual cues and arrangements assists a person in forming and interpreting a mental model (also known as a cognitive map) of some data. If data are encoded so that useful semantic attributes are mapped to strong visual cues, then the resulting representation takes advantage of human perception. A visual feature or visual pattern that attracts a viewer's attention, then, corresponds to an interesting fact or trend in the data. Effective visualization can support needs such as detecting relationships, detecting patterns (both seeing trends and outliers), tracking events over time, change detection/monitoring, and improving situation awareness.

As final points motivating the value of visualization, here are two comments about how not to use visualization. Visualization can be especially helpful for exploratory analysis; for example, some researchers hypothesize that visualization is appropriate when the user does not yet know what questions to ask or what to look for. However, for well-defined tasks, visualization is not a good substitute if the solution can be automatically calculated. Also, not all visualization is helpful; visualization can certainly be poorly designed. Chartjunk [4], exotic 3D layouts, and animations, for example, may be eye candy that seems interesting but that distracts, confuses, or misleads the viewer. An appropriate visualization for a task might be a simple, straight-forward chart that lets the data shine through.

Visualization Tour Of The Internet

Now it is time for some exercises for the reader! To make these visualization definitions and descriptions less academic and more tangible, follow along on this field trip. Each stop in the tour illustrates visualization and interaction techniques, right in your own cyber backyard. Together, the stops were selected to show you a range of possibilities that open data to exploration and that help a viewer understand sometimes subtle relationships in the data.

Exploring Relationships: Visual Thesaurus

Tour Stop: <http://www.visualthesaurus.com/>

Visual Thesaurus is an interactive visualization of word relationships in the English language [5]. The site is a subscription-based website. However, you can test drive the interface on a trial basis. To see the visualization and interaction features, enter the word *quiet*. *Quiet* is a good example word because it has a rich set of associations. (*Good* and *abstract* are also good example words.)

Searching on the word *quiet* results in a view displaying all of the definitions of *quiet* and other words related to those definitions (see Figure 5). Definitions appear as circles and are colored according to their parts of speech (i.e., nouns, adjectives, verbs, and adverbs). Words are connected to their definitions by solid lines. All of the words connected to a single definition are synonyms. Antonyms are linked to a definition by dashed red lines.

Definitions can also link to other definitions by dashed lines that represent a variety of relationships. Visual Thesaurus currently represents 16 different types of relationships, such as “is similar to,” “see also,” and “is a type of.” By hovering the mouse pointer over a line, the specific relationship depicted by the line appears.

At the right side of the display, all definitions in the current view are listed in groups according to the four parts of speech. You can browse the definitions in the current view either by hovering your mouse over the visual side or by moving your mouse over the list entries. Moreover, the two sides are coordinated; highlighting a definition on one side will highlight the corresponding location on the other side (see Figure 6).

The Visual Thesaurus view is oriented around a single focus item (*quiet* in the case of our running example), which is placed in the center of the view. To choose a new focus, left-click on the word or definition (either in the visual side or list side). For certain words and definitions, right-clicking expands the item without refocusing the view. Using the these options, you can journey through the English language in any direction, using the current display as a starting point.

Visual Thesaurus uses the concept of graphs as its visual technique. Here, graph is used, as in computer science and graph theory, to mean a set of relationships between entities. Nodes represent entities, and edges represent the relationships between entities. Renoir is another example of a graph-based visualization tool.

From the standpoint of interface design, Visual Thesaurus successfully tackles two challenges. The first challenge is one of data scale: How does the tool cope when it cannot display all of the user's data in a single visualization? After all, Visual Thesaurus includes over 140,000 words and 115,000 definitions. The second challenge is one of data fusion: Can the tool provide effective, synthesized views of multiple types of data? Visual Thesaurus addresses data scale by providing navigation through subsets of the full graph of the English language. A user views visually manageable, yet still useful, subsets of data and can easily move to new subsets. Visual Thesaurus addresses data fusion by combining different node and edge types within its graphs yet uses color and line styles to visually distinguish the multiple types of data. A user works with all of the data in a unified way.

Animating Scatterplots: Gapminder

Tour Stop: <http://www.gapminder.org/world/>

Gapminder World is a site that displays data (e.g., from the World Bank and UNICEF) about global development trends, for example, the changing relationship between income levels and child mortality [6]. Gapminder World describes its goal as

providing animated statistics to promote a fact-based understanding of our world. To do so, Gapminder uses a scatterplot as its visual technique. Its scatterplots are souped up with flexible, interactive options that include animation.

Using drop-down menus, you can choose the data for the x-axis and y-axis of the scatterplot as well as the data scale (linear or logarithmic). Using the color and

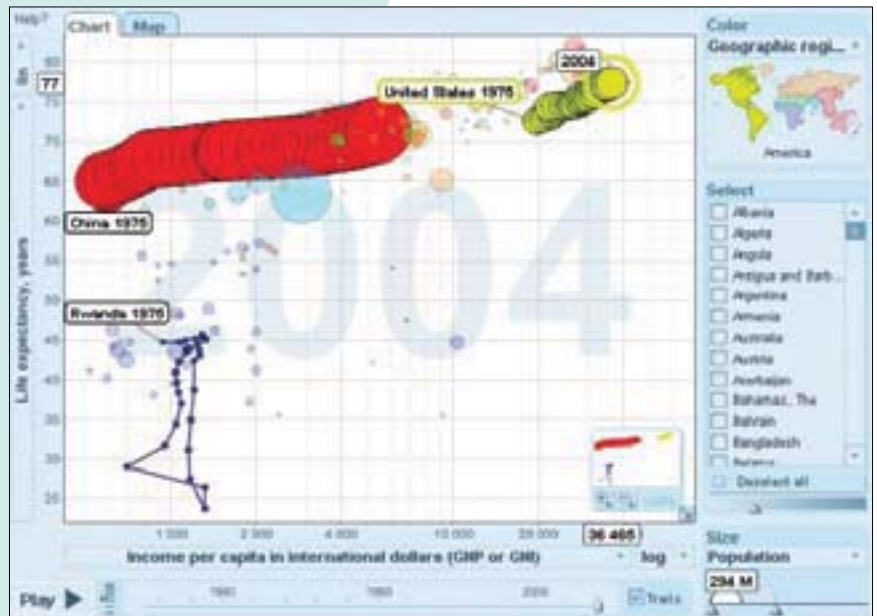


Figure 7: Gapminder World view showing relationship between income level and life expectancy for three countries (time period 1975 to 2004)

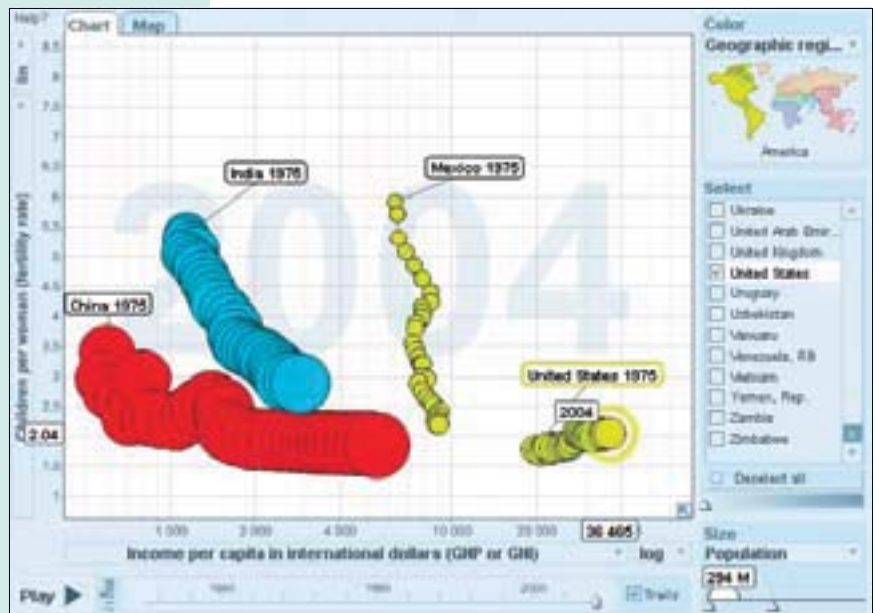


Figure 8: Gapminder World view showing relationship between income level and fertility rate for four countries (time period 1975 to 2004)

Visualization from Gapminder World, powered by Trendalyzer from www.gapminder.org.



Figure 9: Map of the Market, using the blue/yellow color scheme, taken on February 11, 2008

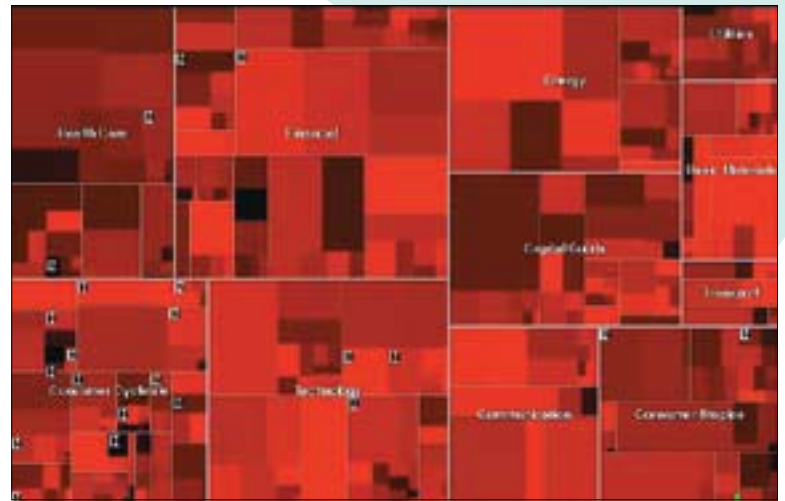


Figure 10: Map of the Market, using the red/green color scheme, taken on February 27, 2007

size panels, you also choose how the data is colored and sized. By moving your mouse over the color legend, you can highlight data points in the scatterplot. You can select items by clicking on points within the scatterplot and by checking items in the selection list. Hovering your mouse over a point in the scatterplot reveals the precise information for that point.

The benefits of the interactive features are magnified when combined with the ability to animate the scatterplot. Animation displays the data over time and, for pre-selected points, creates trails that reveal trends over time. As an example, Figure 7 shows upward trends in life expectancy and income level in the United States and China since 1975, but a very different story, reflecting civil war and genocide, for Rwanda. Figure 8 explores how fertility rates and income levels have changed for the United States, China, India, and Mexico.

From an interface design standpoint, Gapminder shows how simple visual and interactive techniques are powerful in the right mix. Hardly an exotic technique, scatterplots are often helpful for representing numerical data. By combining selection and animation, Gapminder overcomes one difficulty with animation, namely human memory limitations. Whereas people might see an interesting progression while watching an animation, they probably will not remember details after the animation has completed.

Gapminder's animation trails enable a user to notice overall changes in the data while also remembering the specific progression of the data of most interest.

Grouping Data: Stock Market Map and News Map

Tour Stop: <http://www.smartmoney.com/marketmap/>

The SmartMoney Map of the Market provides a view of more than 500 stocks grouped into sectors (e.g., health care, financial, and consumer cyclicals) and industries (within health care, the industries are pharmaceuticals, medical products, biotechnology, and managed care) [7]. Within an industry, the visual layout is fixed so that neighboring companies have historically had similar stock price movements. As a result of this layout, the user has increased ability to notice larger market movements, with trends appearing as regions of light and dark. With daily performance of all of the stocks displayed in a visualization that fits into a single screen, the user immediately gets an overview for the current trading day.

Figure 9 is a map that captured the state of the market on a mixed day, with a fair amount of gainers, but some losers, particularly in the financial sector. Each square represents a single company and is sized by the company's market capitalization. Each square is colored by the company's stock price performance. This example is using a blue/yellow color scheme. Yellow areas indicate stocks with



Figure 11: Control Panel for the SmartMoney Map of the Market

SmartMoney.com ©2008. SmartMoney is a joint publishing venture of DowJones & Co and Hearst SM Partnership. SmartMoney is a registered trademark. All rights reserved. Used with permission from SmartMoney.

price gains and blue areas indicate losses, compared to prices from the previous day's market close. Dark colors indicate little or no price change. Figure 10, a map with a red/green color scheme, shows a bearish day in which the Dow Jones Industrial Average fell 3.29 percent, which was the biggest market drop seen in the prior four years. Losses were across all sectors, one exception being the happy, green company in consumer staples.

Figure 11 shows the Map of the Market control panel, which includes the color encoding being used. Using



Figure 12: Newsmap, taken on February 11, 2008, showing world, national, technology, and entertainment headlines in the United States media

the control panel, you can select the color scheme (blue/yellow or red/green). You can choose to augment the map by marking companies appearing in current news stories or by highlighting the top five stock gainers or losers. You can select how price changes are displayed (e.g., change since yesterday's closing market price or change over the past year). Finally, you can use a search feature to locate individual companies.

Within a map, you can click on companies and headline icons to drill down to details. Clicking a company opens a menu that allows you to zoom into a sector or industry or to access company details. Clicking a headline icon presents the related news article. For example, in Figure 9, a user can inspect one of the bigger price decreases, American International Group, and read the related article "AIG Shares Sink As Credit Losses Mount."

Map of the Market uses a visualization technique called a treemap [8][9]. The treemap design is a technique for showing data in groups, especially for data that forms hierarchical grouping. Data in the same group are displayed together within a contiguous block on the screen.

Hierarchical groups are displayed as nested blocks. In the case of the Map of the Market, the grouping helps a user notice sector trends and exceptions to those trends.

Newsmap is another application of treemaps (Extra Credit Tour Stop: <http://www.marumushi.com/apps/newsmap/newsmap.cfm>) [10]. Newsmap is a visual version of Google's News Aggregator [11]. The Google News Aggregator allows users to search and browse news stories appearing in 4,500 different news sources. Newsmap presents a summary of the headlines (see Figure 12), highlighting the stories receiving the most attention (as measured by the number of articles on the topic) for the indicated time period. The Smithsonian Institute (Extra Credit Tour Stop: <http://historywired.si.edu/index.html>) offers a treemap-based exhibit of 450 museum artifacts.

From the standpoint of interface design, treemaps maximize the use of screen real estate. Compared to a graph layout, which can contain a lot of white, unused space, compact treemaps use all of the pixels to represent data. Treemaps also are an example of a design principle within the information visualization community:

"Overview, zoom & filter, details-on-demand." This mantra, conceived by the eminent visualization researcher Ben Shneiderman, captures the idea of initially providing a user with an overview of a dataset. The overview provides a context for full understanding. From the overview and based on the question at hand, the user can navigate and drill down to the interesting details.

Spotting Trends Over Time: Baby Name Popularity

Tour Stop: <http://www.babynamewizard.com/namevoyager/>

[babynamewizard.com/namevoyager/](http://www.babynamewizard.com/namevoyager/)

NameVoyager is a visual interface to the most popular names for babies born in the United States [12][13]. The database comprises the top 1000 boys and girls names for each decade since the 1890s. The overall NameVoyager view encompasses 5000 names.

The data is maintained and available from the Social Security Administration (SSA). You can visit the SSA website (<http://www.ssa.gov/OACT/babynames/>) for more information about the data collection. Moreover, the SSA website provides a textual comparison to NameVoyager. The webpage <http://www.ssa.gov/OACT/babynames/decades/names1990s.html> shows the list of popular names for the 1990s. NameVoyager includes this decade along with the rest of the century.

Looking at a NameVoyager view (see Figure 13), each strip is a popularity measure for one name. A strip's thickness is based on the name's frequency of use for that decade; the thickness of a strip varies over time. The overall height of the graph shows the cumulative use for the currently displayed names. Names appear in alphabetical order, from top to bottom. Each name is colored as either a girl's or boy's name.

Starting from the initial view of all top-ranked names, you can explore the data by hovering the mouse, clicking on a strip, or doing a prefix search. Mousing over a name will highlight its strip and display its rank for the given decade. Clicking a

name will zoom in on that name. Most powerfully, doing a prefix search will zoom in on a group of names. To conduct a search, click to activate the search box. Type a letter and NameVoyager will zoom into all names beginning with that letter. You can add more letters to your search to narrow the focus of the graph. Notice that, as the graph changes, NameVoyager adjusts the scale of the y-axis to maximize visibility.

Exploring with NameVoyager, you can answer specific questions and see larger trends. After searching for your name (and your spouse's, children's, and parents' names...), you may ask historical and cultural questions (e.g., What has happened to the name Adolph? Are people naming their children Lexus?). In doing prefix searches, you may start to notice group trends. For example, Figure 14 shows that names beginning with "I" are rebounding. Does this pattern hold true for other vowels?

NameVoyager uses a visualization technique known as stack (or stacked) graphs. Each strip, individually equivalent to a single line graph, rests on top of a stack of other line graphs. From the standpoint of interface design, stack graphs were designed to visualize change over time for groups of non-negative quantities, especially if it makes sense to add up the groups. As a counterexample, it generally does not make sense to add together the

stock prices of different companies. For stock prices, line graphs would probably be a better visual choice.

Collaborative Visualization: Many Eyes

Tour Stop: <http://services.alphaworks.ibm.com/manyeyes/home>

Launched in 2007, Many Eyes is a participatory website where visitors can upload data, create interactive visualizations, and contribute comments [14][15]. By providing a website and visualization tools, its developers hope to spark collaborative, deliberative analysis and discussion about data.

A Many Eyes user can view and comment on data and visualizations created by others. A user can also load new data and create visualizations based on how the data is formatted. Currently, Many Eyes provides roughly fifteen interactive visualization options, including treemaps, stack graphs, and bubble charts. Figure 15 shows a bubble chart of data from the US Centers for Disease Control and Prevention (CDC) of gastrointestinal illness outbreaks on international cruise ships. Each bubble represents a cruise ship line, and colored slices within a bubble represent the number of ill passengers for each ship in the cruise line. Visualizations on Many Eyes can be collected into topic hubs to allow people with specific interests to focus attention on relevant data. For example, Many Eyes

currently has relatively active topic hubs on food safety issues, crime statistics for the United States, terrorism, and presidential campaigns.

As a social experiment, Many Eyes looks at the question of what happens when large numbers of people can access visualization technology. Many Eyes exemplifies a trend in the information visualization community towards giving people greater access to data and visual analysis tools. Many Eyes also exemplifies the concept of Web 2.0, which is the idea of using the Internet as the platform for information sharing and social collaboration. Proponents of Web 2.0 see the World Wide Web as a place to maximize individual and collective intelligence by enabling users to contribute as well as to consume information.

Visualization for the Masses

The trend towards giving people greater access to data and useful visualizations is also occurring in the press. Both print and web-based publications are using visualizations as access to data. Consequently, readers have a chance to assess whether the data support conclusions presented in the media.

Tour Stop: <http://www.slate.com/id/2180392/>

Slate, a daily web-based magazine, provides an example of using a social network graph (see Figure 16)[16].

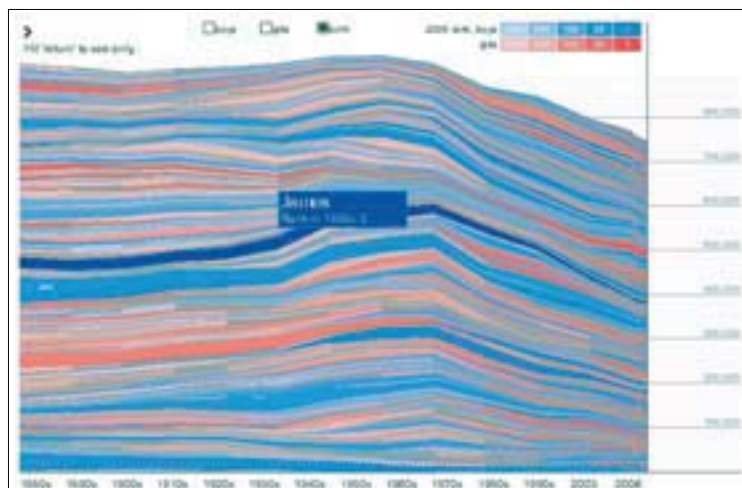


Figure 13: NameVoyager interface highlighting the long-term popularity of the name James

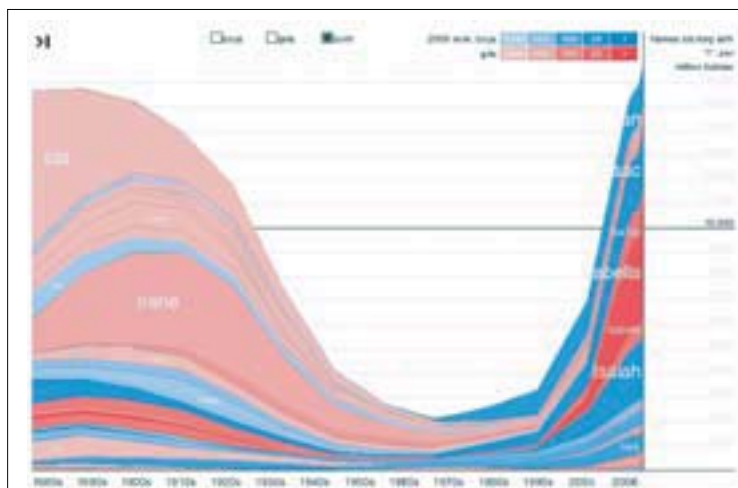


Figure 14: NameVoyager interface showing a resurgence of interest in names beginning with the letter "I"

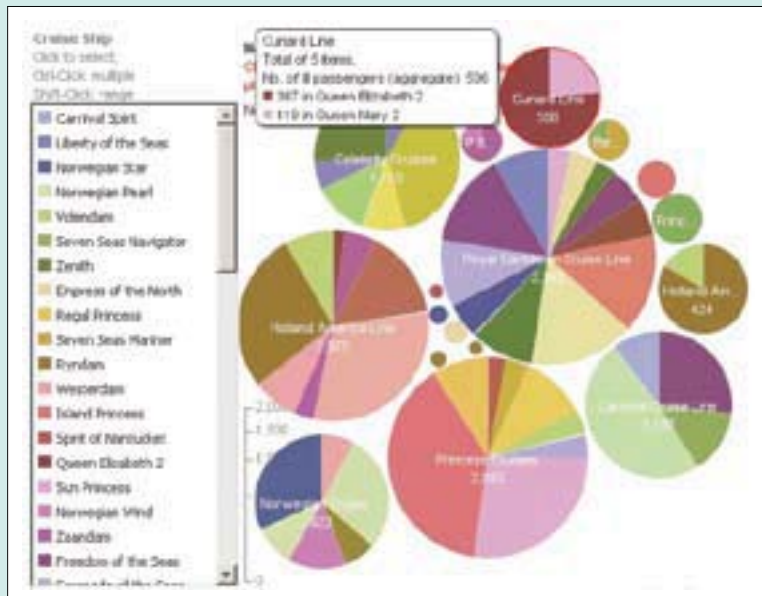


Figure 15: Many Eyes bubble chart titled “Outbreak Summaries for International Cruise Ships: Jan. 2004 - Aug. 2007”

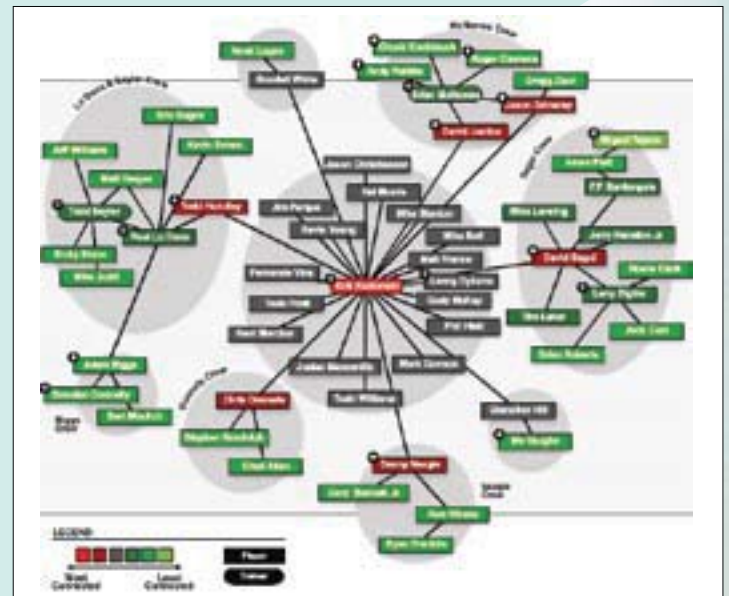


Figure 16: Slate's summary of the Mitchell Report

The interactive graph summarizes the Mitchell Report about performance-enhancing drugs in professional baseball. The graph condenses the 409-page report into key findings about people and their relationships. For a reader who does not also read the full Mitchell Report, the graph serves as a comprehensible summary of the material. For a reader who also reads the full report, the graph helps the reader organize, remember, and review the findings.

The New York Times provides many other examples of static, in-print graphics and dynamic, on-line visualizations. Matt Ericson, deputy graphics editor at The New York Times, recently spoke about how the information graphics team produces visualizations for a broad audience [17]. Using the principle “Show, Don’t Tell,” the team strives to present raw data clearly so that a reader does not have to comb through an associated article to understand the visualization. They pay close attention to accurate portrayal, avoiding visualizations that tend to cause mistaken conclusions. They augment their visualizations with textual descriptions to guide readers to interesting areas of the data. They provide more than one view when needed to explain a complex or subtle point. In this

way, Ericson emphasized that the team members function as journalists rather than illustrators.

Tour Stop: http://www.nytimes.com/interactive/2007/10/23/us/20071023_FIRES_GRAPHIC.html#

Tour Stop: http://www.nytimes.com/ref/sports/20070731_BONDS_GRAPHIC.html#

Tour Stop: http://www.nytimes.com/ref/washington/20070123_STATEOFUNION.html

Tour Stop: http://www.nytimes.com/interactive/2007/12/13/us/politics/20071213_DEBATE_GRAPHIC.html#transcript

As an example, The New York Times published a visualization of the wildfires in southern California in October 2007 [18]. The online visualization integrated data from California state and county emergency management agencies and the United States Forest Service, providing a direct understanding of the extent and progression of the fires. An example of sports data, a visualization of Barry Bonds home runs record shows the accumulation of home runs for the 21 professional players with the most home runs in baseball [19]. Looking at Barry Bonds’

performance from shortly after home run 400 to home run 600, Bonds’ rate of home runs increases (reflected by the slope of the line). This acceleration differs from other players, whose home run rates tend to plateau or decrease when they reach their thirties. According to claims about steroid use, Bonds allegedly began taking steroids before his 14th season (shortly after his 400th home run). Final examples from The New York Times focus on the words that politicians use in speeches and how their messages change over time [20][21]. Views of State of The Union speeches (using bubble charts) and presidential candidate debates (via a transcript analyzer) enable readers to find passages relevant to their interests (e.g., budget, social security, or war).

References

- [1] Card S, Mackinlay J, & Shneiderman B. Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann Publishers; 1999.
- [2] Thomas JJ & Cook K, editors. Illuminating the Path: The Research and Development Agenda for Visual Analytics [Internet]. IEEE Computer Society; 2005 [retrieved 26 Feb 2008]. Book available at: <http://nvac.pnl.gov/agenda.stm>

- [3] Bremner E & Fellenz C. The Price of Power. Baltimore Sun. 11 June 2006.
- [4] Tufte ER. The Visual Display of Quantitative Information. Graphics Press; 1983.
- [5] VisualThesaurus[Internet].Thinkmap, Inc; 1998-2008 [retrieved 11 Feb 2008]. Available at: www.visualthesaurus.com
- [6] Gapminder [Internet]. Gapminder World [retrieved 11 Feb 2008]. Available at: <http://www.gapminder.org/world>
- [7] SmartMoney Map of the Market [Internet]. Dow Jones & Company, Inc. and Hearst SM [retrieved 11 Feb 2008 and 27 Feb 2007]. Available at: <http://www.smartmoney.com/marketmap/>
- [8] Johnson B & Shneiderman B. Tree-maps: A Space-Filling Approach to the Visualization of Hierarchal Information Structures. In: Proceedings IEEE Information Visualization 1991; p. 297-282.
- [9] Shneiderman B. Treemaps for Space-Constrained Visualization of Hierarchies [Internet]. [retrieved 25 Feb 2008]. Available at: <http://www.cs.umd.edu/hcil/treemap-history/>
- [10] Newsmap [Internet]. Weskamp M & Albritton D [retrieved 11 Feb 2008]. Available at: <http://www.marumushi.com/apps/newsmap/newsmap.cfm>
- [11] Google News Aggregator [Internet]. [retrieved 11 Feb 2008]. Available at: <http://news.google.com/>
- [12] Baby Name Wizard's Name Voyager [Internet]. Generation Grownup LLC [retrieved 11 Feb 2008]. Available at: <http://www.babynamewizard.com/namevoyager/>.
- [13] Wattenberg M. Baby Names, Visualization, and Social Data Analysis. In: Proceedings IEEE Information Visualization 2005; p 1-7.
- [14] Many Eyes [Internet]. Visual Communication Lab, IBM [retrieved 15 Feb 2008]. Available at: <http://services.alphaworks.ibm.com/manyeyes/home>
- [15] Viegas F, Wattenberg M, van Ham F, Kriss J, & McKeon M. Many Eyes: A Site for Visualization at Internet Scale. IEEE Transactions on Visualization and Computer Graphics. 2007;13(6):1121-1128.
- [16] Perer A & Wilson C. The Steroids Social Network: An Interactive Feature on the Mitchell Report. Slate [Internet]. updated 21 Dec 2007 [retrieved 14 Feb 2008]. Available at: <http://www.slate.com/id/2180392/>
- [17] Ericson M. InfoVis Keynote Address. Visualizing Data for the Masses: Information Graphics at The New York Times. IEEE Transactions on Visualization and Computer Graphics. 2007;13(6):xxv. Additional information available at <http://www.ericson.net/>, [retrieved 22 Feb 2008]
- [18] Aigner E, Bloch M, Carter S, Cox A, & Ericson M. The Extent of the Fires. The New York Times [Internet]. 23 Oct 2007 [retrieved 26 Feb 2008]. Available at: http://www.nytimes.com/interactive/2007/10/23/us/20071023_FIRES_GRAPHIC.html
- [19] Carter S, Cox A, & Ward J. Paths to the Top of the Home Run Charts. The New York Times [Internet]. 08 Aug 2007 [retrieved 26 Feb 2008]. Available at: http://www.nytimes.com/ref/sports/20070731_BONDS_GRAPHIC.html#
- [20] Werschkul B. The 2007 State of the Union Address. The New York Times [Internet]. [retrieved 26 Feb 2008]. Available at: http://www.nytimes.com/ref/washington/20070123_STATEOFUNION.html
- [21] Carter S, Dance G, Ericson M, Jackson T, Singh V, Ellis J, & Sharpe M. Democratic Debate: Analyzing the Details. The New York Times [Internet]. 13 Dec 2007 [retrieved 26 Feb 2008]. Available at: http://www.nytimes.com/interactive/2007/12/13/us/politics/20071213_DEBATE_GRAPHIC.html#transcript



Visual Analytics: Adventures in Exploring Information Spaces

On a recent adventure in a remote part of the Grand Canyon, I gazed at a panel of rock art depicting the revered pursuit of an ancient hunt. These 2,000 year-old petroglyphs¹ seemed to tell the story of a hunter and the hunted. A man armed with a bow and arrow was tracking a bighorn sheep. At least, that's the interpretation that unfolded in my mind. In reality, the meaning and purpose of such drawings is unknown and is often debated by scholars. These early static visualizations chiseled in rock have outlasted both time and meaning. We do not know the original intention of the ancient artist. Was he telling a story, recording history, paying homage to the gift of food, relaying info to his buddies where to find good hunting? It's a mystery, but clearly, he was recording some important notion for himself and others. We don't know what he was thinking or intending as he rendered his masterpiece.

A static drawing or visualization with limited additional supporting information is easily misinterpreted. Its meaning can be difficult to share with others. A good, useful, well-conceived visualization may readily show relationships or concepts that answer specific questions. However, what if you want to ask a slightly different question? What if you want to vary one parameter? What if you want to explore an idea that was not anticipated by the original “artist”? You don’t want to gaze at a static piece of “rock art” and guess an interpretation. You, the curious, delving intelligence analyst, want to explore many facets of your information space. You want to look at data from multiple perspectives and test different theories. You want to examine, compare, and contrast multiple relationships and patterns. You require a dynamic, interactive, visual exploration environment to conduct your own adventures in analysis. Providing such a useful analytic exploratorium is exactly the goal of a new multi-disciplinary field of research known as visual analytics. Visual Analytics focuses on injecting interactive visualization throughout the analysis process. Here visualization is not an after-thought or used only for presentation but an integral part of analysis.

This article is an introduction to the new, exciting research area of visual analytics. We outline the multi-disciplinary nature of this work and point to its roots in supporting fields. We then focus on a new national research initiative in visual analytics known as the National Visualization and Analytics Center (NVAC). NVAC is sponsored by the Department of Homeland Security (DHS) and implements a broad research agenda focused on homeland security applications in academia, industry, and government. The National Security Agency (NSA) is an active participant in these NVAC efforts. We briefly discuss our relationship and focus on a current research challenge problem that NVAC is pursuing for the Research Directorate’s Knowledge Discovery Research group. In the process, we mention a few examples of in-house ef-

forts for applying ideas of visual analytics to intelligence problems. An over-arching goal of this article is to inform and inspire you to think of your problems, analyses, and information spaces in terms of visual analytics. Are you stumped, gazing at your own “rock art”? Perhaps, the NVAC program can help you become a full-fledged “archeologist” in your analytic pursuits.

Visual Analytics

Everything old is new again.

There is nothing new under the sun.

Visual Analytics is often touted as a new emerging discipline that combines efforts from a variety of other fields. This is true. However, in many respects, it might also be viewed as the maturing of visualization applied to the complete spectrum of analysis. One could argue that the very earliest computer graphics such as Ivan Sutherland’s Sketchpad system were also early forms of visual analytics [1]. However, in recent years there has been a key confluence of events: 1) the maturing of dependent sub-fields of research; 2) sufficiently powerful, affordable, interactive computers and software; 3) scale: the explosion in the sizes and complexity of datasets that outstrip traditional analysis methods; and 4) wide-spread demand for better analysis methods to deal with this complexity by many people from many diverse domains. The combination of these new technological capabilities and experiences as well as the pent-up demand for better analysis methods seems to have coalesced into a critical nexus.

The oft-cited birth of this new push in modern visual analytics is traced to a research panel gathered by DHS following the September 11 attacks on the United States. The subsequent response to these horrific events by the US Government took many forms. The establishment of DHS itself was one such result. One general thrust within DHS was to develop new technologies to provide technical and analytical advantages in preventing terrorism. With the increased complexities of analysis, technologies were needed to bet-

ter support human judgment and analysis. The intent was to make the best possible use of intelligence information and appropriately share it with others to prevent, deter, and respond to threats. Therefore, in 2004, DHS chartered and established NVAC specifically to address the need for better analysis through visualization [2].

A major objective of NVAC was to define a long-term research and development agenda for visual analytics to facilitate advanced analytical insight. Based in part on a long, well-established record of laudable research in information visualization, DHS chose Pacific Northwest National Laboratory (PNNL) to oversee and lead the NVAC effort. NVAC pulled together a multi-disciplinary panel of top researchers from diverse fields to develop this research agenda through a series of workshops. The panel’s resulting research was published as a well-received and comprehensive 2005 publication titled “Illuminating the Path: The Research and Development Agenda for Visual Analytics” (see Figure 1) [3,4,5]. This noteworthy publication is a key milestone in launching this new field of visual analytics as well as providing a long-term research agenda to achieve its goals. It focuses on the need for collaborative and cooperative work in academia, government, and industry to improve analysis in Homeland Security applications.

In this publication, the panel carefully chose to define visual analytics as “the sci-

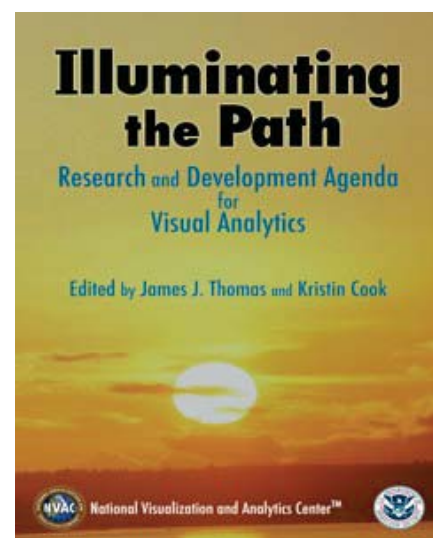


Figure 1: NVAC report: Illuminating the Path

¹ Petroglyphs are images chiseled in rock. Pictographs are images painted on rock.

ence of analytical reasoning facilitated by interactive visual interfaces.”

Visual Analytics is the science of analytical reasoning facilitated by interactive visual interfaces.

Visual Analytics as a Research Discipline

One very important factor that differentiates visual analytics from its supporting fields of information visualization and human-computer interaction (HCI) is its emphasis on analytical reasoning. Traditional visualization focuses on visual representations and visual mappings of data. HCI focuses on effective interaction in terms of usability and utility. However, how these interactive visualizations best fit into the analytic cognitive workflow has often been ignored. The act of using visualization often takes analysts out of their analytic “zone” or cognitive working model. Visual analytics emphasizes analytical reasoning and attempts to blend visualization throughout the analytic process without violating this cognitive workflow. An entire early section of the NVAC report (chapter 2) focuses on “The Science of Analytical Reasoning” and discusses visualization’s role in the analytic sense-making loop.

The NVAC report also characterizes the analytic use of visual analytics in four main ways:

- Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data.
- Detect the expected and discover the unexpected.
- Provide timely, defensible, and understandable assessments.
- Communicate assessment effectively for action.

Since analysis is very much a multi-disciplinary endeavor, visual analytics is also multi-disciplinary in nature. It draws heavily from a wide range of fields including information visualization, interaction, cognitive science, knowledge discovery, computer science, math, statistics, perception, and data management, as well as others (see Figure 2). One of the main

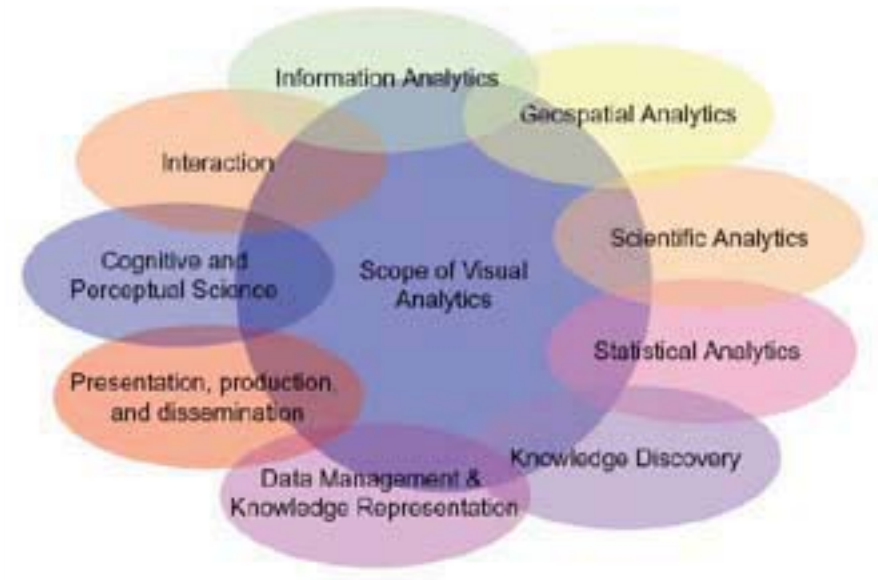


Figure 2: The Multi-disciplinary Scope of Visual Analytics

goals of visual analytics is better integration and fertilization of these disciplines. Hence, the NVAC report strongly suggests multi-disciplinary education and training in some of these key areas. One thrust of NVAC is to educate and train the next generation of researchers and practitioners in visual analytics. Hence NVAC is working with academia, industry, and government to help formulate and plan a visual analytics curriculum.

National Visualization and Analytics Center

NVAC is a multitude of things. It’s an idea. It’s a DHS research program. It’s a government resource, an academic resource, and an industry resource. It’s a network of collaborative researchers and centers. It’s an education resource, an analytic resource, and technical resource. In addition to all of these virtual facets, the NVAC program is also a real, physical center of excellence. NVAC is located at PNNL, in Richland, WA. When DHS decided to form NVAC, they chose PNNL to direct and run the NVAC program based on PNNL’s previous experience and research in visualization and government analytic applications. Dr. Jim Thomas of PNNL was selected to be director of NVAC and serve as the nation’s visual analytics champion.

From the very beginning, Dr. Thomas envisioned a network of key academic re-

search centers where NVAC would guide (and fund) research and education in visual analytics. These Regional Visualization and Analytics Centers (RVACs) would be primary links to academia and research and serve as regional centers of excellence. Government and industry partners could work directly with RVACs and help drive their research problems. An RVAC can be a single institution or a team of institutions. Specific RVACs might focus on particular domains or research areas. Similarly, Dr. Thomas envisioned Government Visualization and Analytic Centers (GVACs). These government centers would help drive NVAC analytic problems as well as serve as test beds and technology delivery conduits for NVAC developed capabilities and technologies.

In addition to directing and funding research at these RVACs, NVAC itself conducts research at its PNNL facilities. To guide this research, NVAC seeks analytic input and challenge problems from the intelligence community, DHS, and other parts of government. To date, two government centers (GVACs) have been established and actively participate in the NVAC research program: CIA and NSA. The Canadian government has also followed the development of NVAC with much interest. Just recently, the Defence R&D Canada announced their own NVAC-like research program, nominally

- [3] Bremner E & Fellenz C. The Price of Power. Baltimore Sun. 11 June 2006.
- [4] Tufte ER. The Visual Display of Quantitative Information. Graphics Press; 1983.
- [5] VisualThesaurus[Internet].Thinkmap, Inc; 1998-2008 [retrieved 11 Feb 2008]. Available at: www.visualthesaurus.com
- [6] Gapminder [Internet]. Gapminder World [retrieved 11 Feb 2008]. Available at: <http://www.gapminder.org/world>
- [7] SmartMoney Map of the Market [Internet]. Dow Jones & Company, Inc. and Hearst SM [retrieved 11 Feb 2008 and 27 Feb 2007]. Available at: <http://www.smartmoney.com/marketmap/>
- [8] Johnson B & Shneiderman B. Tree-maps: A Space-Filling Approach to the Visualization of Hierarchal Information Structures. In: Proceedings IEEE Information Visualization 1991; p. 297-282.
- [9] Shneiderman B. Treemaps for Space-Constrained Visualization of Hierarchies [Internet]. [retrieved 25 Feb 2008]. Available at: <http://www.cs.umd.edu/hcil/treemap-history/>
- [10] Newsmap [Internet]. Weskamp M & Albritton D [retrieved 11 Feb 2008]. Available at: <http://www.marumushi.com/apps/newsmap/newsmap.cfm>
- [11] Google News Aggregator [Internet]. [retrieved 11 Feb 2008]. Available at: <http://news.google.com/>
- [12] Baby Name Wizard's Name Voyager [Internet]. Generation Grownup LLC [retrieved 11 Feb 2008]. Available at: <http://www.babynamewizard.com/namevoyager/>.
- [13] Wattenberg M. Baby Names, Visualization, and Social Data Analysis. In: Proceedings IEEE Information Visualization 2005; p 1-7.
- [14] Many Eyes [Internet]. Visual Communication Lab, IBM [retrieved 15 Feb 2008]. Available at: <http://services.alphaworks.ibm.com/manyeyes/home>
- [15] Viegas F, Wattenberg M, van Ham F, Kriss J, & McKeon M. Many Eyes: A Site for Visualization at Internet Scale. IEEE Transactions on Visualization and Computer Graphics. 2007;13(6):1121-1128.
- [16] Perer A & Wilson C. The Steroids Social Network: An Interactive Feature on the Mitchell Report. Slate [Internet]. updated 21 Dec 2007 [retrieved 14 Feb 2008]. Available at: <http://www.slate.com/id/2180392/>
- [17] Ericson M. InfoVis Keynote Address. Visualizing Data for the Masses: Information Graphics at The New York Times. IEEE Transactions on Visualization and Computer Graphics. 2007;13(6):xxv. Additional information available at <http://www.ericson.net/>, [retrieved 22 Feb 2008]
- [18] Aigner E, Bloch M, Carter S, Cox A, & Ericson M. The Extent of the Fires. The New York Times [Internet]. 23 Oct 2007 [retrieved 26 Feb 2008]. Available at: http://www.nytimes.com/interactive/2007/10/23/us/20071023_FIRES_GRAPHIC.html
- [19] Carter S, Cox A, & Ward J. Paths to the Top of the Home Run Charts. The New York Times [Internet]. 08 Aug 2007 [retrieved 26 Feb 2008]. Available at: http://www.nytimes.com/ref/sports/20070731_BONDS_GRAPHIC.html#
- [20] Werschkul B. The 2007 State of the Union Address. The New York Times [Internet]. [retrieved 26 Feb 2008]. Available at: http://www.nytimes.com/ref/washington/20070123_STATEOFUNION.html
- [21] Carter S, Dance G, Ericson M, Jackson T, Singh V, Ellis J, & Sharpe M. Democratic Debate: Analyzing the Details. The New York Times [Internet]. 13 Dec 2007 [retrieved 26 Feb 2008]. Available at: http://www.nytimes.com/interactive/2007/12/13/us/politics/20071213_DEBATE_GRAPHIC.html#transcript

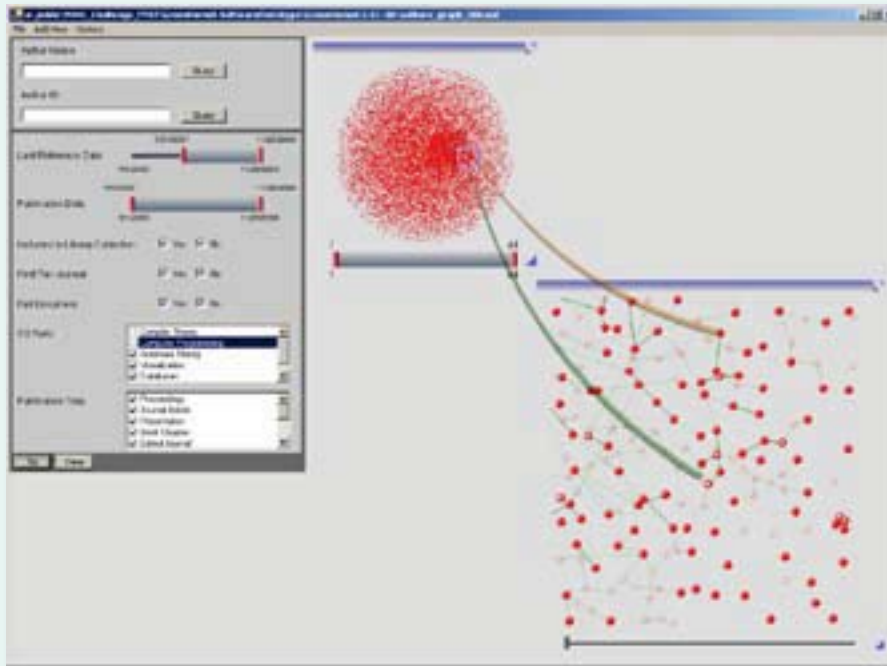
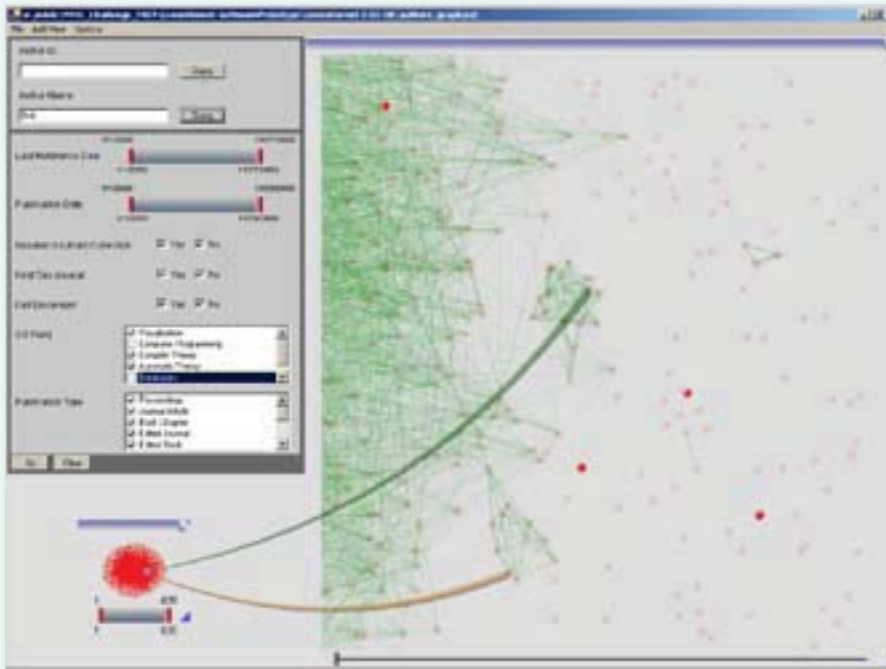


Figure 3: Interactive Visualization and Navigation of Large-Scale Graphs

Figure 4: Multiple Level- ℓ Detail and Abstraction in Large-Scale Graphs

filter and details-on-demand paradigm is much more effective for understanding large datasets than a bottom-up query and display paradigm; 3) Interactive visual discovery methods are necessary for finding the “unknown unknowns”;² 4) Size and complexity of information spaces have outpaced the limits of many visualization capabilities.

In this work, progressive disclosure and multiple levels of detail are used to

abstract parts of the graph. An overview window is used to maintain overall context of the entire dataset. Figures 3 and 4 are examples and early results of this NVAC research challenge problem using a synthetic dataset. Our synthetic test dataset consists of a co-author network of paper publications in which a node represents an author and a link represents a co-authored paper. It should be noted that *interactive* display and navigation of graphs of this


size is an order of magnitude improvement beyond most cited and recognized capabilities. This has truly been a groundbreaking effort. For the coming year, we are continuing this work with new emphasis on using analytics to guide and abstract visual representation and layout during navigation and interaction. We consider this initial prototype to be an interesting base capability for further testing and exploring a variety of interesting analytics. In summary, we have found participating with NVAC to be very beneficial in supporting our programs in visual analytics.

Conclusion

This article has been a very brief overview and introduction to the new multidisciplinary field of visual analytics. A key focus of visual analytics is the science of analytical reasoning by interactive visual interfaces. This requires tight integration of analytics, visualization, and interaction to provide a rich cognitive exploratorium. NVAC is leading a national research initiative for further research and education in this field with the goal of advancing analysis for homeland security applications. NSA is actively participating in NVAC and is currently funding a challenge problem to advance the state of large-scale graph analysis for analytic applications. We welcome further discussion and participation in our efforts. We also encourage you to learn more about NVAC and participate as appropriate.

Visual Analytics is intended to be an active, engaging exploratory process of discovery. Interactive probing of relationships and testing of hypotheses is facilitated by visual feedback. Reflecting on my amateur interpretation of static Grand Canyon rock art, it is clear that meaningful analysis or interpretation is difficult without further context or supporting information. These 2,000-year-old static visualizations are frozen in time with obscure meaning. Hopefully, as you explore your own information spaces, you will have sufficient analytic tools and a supportive environment to conduct a full “archeologi-

² Finding “unknown unknowns” - identifying the gaps or missing information in an analytic information space, realizing what you don’t know

cal excavation.” The emerging field of visual analytics might provide you with the proper analytic environment to help you turn your static “rock art” into a lucrative mine of gold. 

References and Resources

[1] Sutherland IE. Sketchpad: A Man-Machine Graphical Communication System. Proceedings of the Spring Joint Computer Conference. Baltimore (MD): Spartan Books, 1963.

[2] National Visualization and Analytics Center (NVAC): <http://nvac.pnl.gov>

[3] Thomas JJ & Cook K, editors. Illuminating the Path: The Research and Development Agenda for Visual Analytics [Print]. IEEE Computer Society; 2005. ISBN 0-7695-2323-4.

[4] Thomas JJ & Cook K, editors. Illuminating the Path: The Research and Development Agenda for Visual Analytics [Internet]. IEEE Computer Society; 2005. Book available from <http://nvac.pnl.gov/agenda.stm>

[5] Thomas JJ & Cook K, editors. Illuminating the Path: The Research and Development Agenda for Visual Analytics [Internet movie]. Available from: nvac.pnl.gov/agenda.stm#movie

[6] Regional Visualization and Analytics Centers (RVACs): nvac.pnl.gov/centers.stm

[7] North-East Visualization and Analytics Center (NEVAC): www.geovista.psu.edu/NEVAC/

[8] Purdue Visualization and Analytics Center (PURVAC): engineering.purdue.edu/PURVAC/

[9] Stanford University RVAC: graphics.stanford.edu/~hanrahan

[10] South-East Regional Visualization and Analytics Center (SRVAC): srvac.uncc.edu

[11] Pacific Rim Visualization and Analytics Center (PARVAC): www.hitl.washington.edu/parvac/

[12] IEEE VAST Challenge: www.cs.umd.edu/hcil/VASTchallenge08/

[13] Information Visualization Benchmarks Repository: www.cs.umd.edu/hcil/infovisRepository

[14] Visual Analytics Digital Library: vadl.cc.gatech.edu



Do You Know What's Happening — Just Beyond Your View?

Analysts are always looking for more information to broaden their understanding of a subject in order to provide the best possible assessment. They want to know what is not yet known, to see beyond their current view, and it is this desire that drives the collection of more and more data. This need to understand what is outside the current view is also important when designing representations that support analysts who must work through these resultant large data sets.

In a world where technology enables the processing and analysis of a huge amount of electronic data, it falls to the intelligence analyst to make sense of it all. One area in which technology has attempted to support intelligence analysis is through data visualizations and information displays. Visualizations, in which data are displayed in recognizable graphic elements, are increasingly moving into mainstream applications as a remedy for information overload. Just as technological advancements have improved data collection, attempts have been made

to provide new tools for presenting those data to better support the analytic process. As analysts find growing volumes of data at their disposal, software developers are retuning display perspectives to make the information available faster and more in synch with the requirements of intelligence analysis. In this article, we describe how longshot displays are a critical element to complement the current trend toward powerful but highly focused visualization techniques. Longshot displays explicitly support the analyst's ability to understand data in its larger context, focus and refocus on individual elements of that data set for detailed inspection in traditional visualizations, and prevent the "keyhole effects" that contribute to "Premature Closure" of analytic conclusions [1].

A technology common in many analytic tools is the ability to query a database

to return a data subset that becomes the basis for the day's work. Within this list, the analyst is expected to work through each of the items, select which data to work (typically done in a sequential fashion), and then conduct the appropriate actions. But does this Filter-Select-Analyze (FSA) model accurately represent and completely support the cognitive work that needs to be accomplished by an intelligence analyst? What are the hidden assumptions contained within an FSA model?

The first assumption the FSA model makes is that the intelligence analysts can accurately construct a query that will appropriately filter the database down into a meaningful subset of data that is ready for analysis. When given only the selected results from a query, the analyst is left wondering—what data did I miss and why? After all, analysis is an iterative narrow-

Longshot displays are used as a visualization technique to maintain information awareness or context in a domain display with large amounts of data. The name "Longshot Display" is used to describe a variant of any number of displays that range from a simple scatterplot to more complex applications.

ing and broadening process. In addition to narrowing functions, the support tools analysts use should cater to broadening searches to help recognize relevant information, break fixations on single or brittle hypotheses, and/or widen the hypothesis set that is considered to explain the available data [1].

A second common assumption of visualization techniques is that if the underlying data (e.g., the results from a query) can be represented, then the visualization has done all that it needs to do. The data can be selected from the visualization, therefore it is a good visualization, and the work to understand the data falls on the user. However, simply assuming application programming interface (API) data compatibility between data and presentation layers does not ensure that the analysts are effectively supported. What is necessary is a set of visualizations and technology that supports all of the cognitive work the intelligence analyst must accomplish [2]. Effective support for intelligence analysts requires that visualizations provide mechanisms for supporting a large number of cognitively difficult tasks, including (but certainly not limited to):

- pulling records and running database searches,
- understanding and managing the results of the data search,
- staying on top of new developments while doing research and analysis,
- determining which datum out of the set to select for analysis,
- completing analysis of selected data, and
- determining what kind of information is helpful and what is not.

While no one component technology can facilitate all of the necessary cognitive tasks, the suite of support tools available to the analyst must facilitate the synchronized accomplishment of these tasks. Most visualizations focus only on supporting analysts to run database searches and perform analysis on the returned data, yet they fail to support the other forms of cognitive work. Coupling these detail-focused displays with an information space, longshot displays will provide support for these other forms of cognitive work, enabling analysts to more effectively work with the overall data set. Longshot displays present

all data related to a particular task to an analyst simultaneously in a single global context, providing the proverbial 40-thousand foot view from which the entire data set can be seen and understood.

Characteristics of Longshot Displays

Showing all the data in the data set is a necessary but not sufficient requirement to create a longshot display. An effective longshot display will also have frames of reference (the variables used on the information space axes) that support the information relationship requirements underlying the user's relevant cognitive work. For example, a map of the world with data overlaid on it will show all data on it (since the world will contain all the data), but may not qualify as a longshot display if the analyst does not need the data in a spatial frame of reference provided by the map to complete an analysis on the data.

Establishing the proper frames of reference for the display allows analysts to detect patterns within the data set. In such emergent pattern displays, although each datum is represented, the data as a whole accumulate into larger coherent structures, providing meaning to the data set as a whole.

This design approach is grounded in the fundamental competency of the human perceptual system. Research on human perceptual capabilities has repeatedly shown that our perceptual systems exhibit the tendency to seek out patterns in the world, requiring little cognitive effort to see and understand these patterns. These pattern recognition capabilities are so prevalent that our perceptual systems may even create false patterns when objects have no relationship to one another (e.g., consider how easy it is to spot animals or other objects in cloud formations¹).

When displays are designed to take advantage of our visual system's powerful pattern recognition processing, the analyst doesn't need to remember the value of each datum; the pattern reveals the data set (see: Longshot Example 1). At the same time, as each individual data value is

Criteria for Longshot Displays

In order for a display to be considered a longshot display it must pass several criteria:

- 1) The display must contain all data related to the user's defined cognitive work.
- 2) The data must be placed in a meaningful frame of reference.
- 3) The display must provide mechanisms to support user focusing and refocusing.
- 4) The display must be coupled with a details view, in which users can focus on selected data.

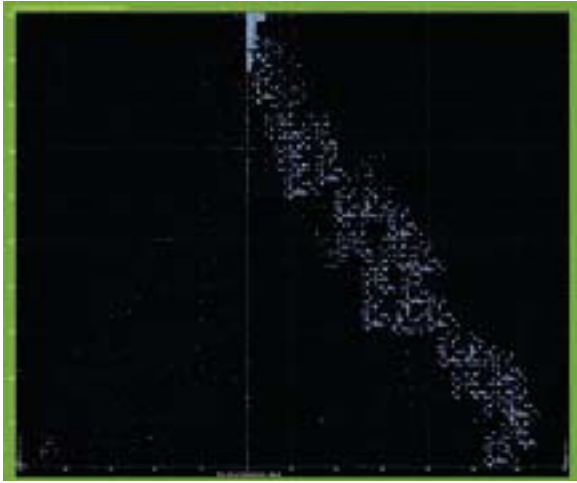
represented on the display, the user is also able to see and determine the value of an individual datum. The analyst is given the best of both worlds.

Text lists and spreadsheets do not support (and can often undermine) this ability to identify patterns and force analysts to think about data one element at a time. Such serial processing is constrained by the limits of human memory and requires significant cognitive processing to make the connections hidden by such "flattened" representations of the data. By presenting the data to take advantage of the ability to see patterns, the representation designer alters how analysts process data, shifting the work from effort intensive cognitive processing into the faster, unconscious, less error prone processing of the visual system. This is equivalent to a computer, in which graphics processing is faster and more efficient when visual information is stored and processed on specialized video hardware instead of using the general purpose CPU.

But how does a representation designer determine what the proper frames of reference should be for the longshot display? The key to answering this question lies in understanding the work the user is trying to accomplish. The cognitive work of the analyst will drive the information relationships needed to complete that work, which will guide the frames of reference to be used. For example, if an analyst's report is due at a certain time, the designer may consider using time until the report is due as a key frame of reference in

¹ Of course, designers must also be careful about creating false patterns with their data representations.

Longshot Example 1: Intelligence Questions Display



An Intelligence Question is a request for information that can be directly answered based on collected evidence in the world. These are de-aggregations of Intelligence Needs, which ask higher level, operational questions.

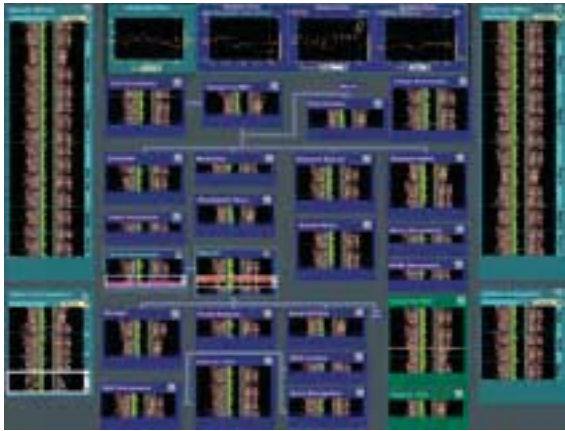
NSA is tracking a large number of Intelligence Questions at a time (the number is large enough that a precise count is not known). Managing that many Intelligence Questions (or any entity, for that matter) is a daunting challenge, but managing these Intelligence Questions is essential to operational success. Therefore, the NSA's Laboratory for Creative Development (LCD) devised a conceptual solution that utilized a Longshot display.

LCD identified the key frames of reference for monitoring the set of Intelligence Questions: the priority of each Intelligence Question, whether reports have been generated in the past on each Intelligence Question, and how long each Intelligence Question has gone without a report. The LCD believed that, while more information existed for each Intelligence Question, these three parameters would allow analysts to decide where in the set of Intelligence Questions to focus.

Those Intelligence Questions that are higher priority are higher on the y-axis of a scatter-plot. On the left side of the y-axis are those Intelligence Questions that are new and have never been reported on, those to the right have been reported on in the past, but require continued scrutiny. Intelligence Questions that are older are further out on the x-axis (making the display reflective along the y-axis). Given that many Intelligence Questions may have the same set value for each parameter, the size of a mark in the visualizations indicates how many Intelligence Questions share those parameters.

The display clearly reveals to the analyst high priority Intelligence Questions that have gone a long time without an assessment (upper corners of the display), as well as low priority Intelligence Questions that have been assessed recently (lower-center of the display). It is easy to see when low priority work is being done over high priority work (although why is left for the details display—an assessment could have been made on a low priority Intelligence Question because the information required was easily available, while not data were available for the high priority Intelligence Questions). The visualization guides analysts to where to focus next, allowing them to better manage the set of thousands.

Longshot Example 2: Performance View



Performance View is a replacement visualization design that supports workers monitoring the health and flow of a processing network. The previous version of the system evolved as an aggregation of individual sensor display elements, reaching a real estate constraint that at its upper limit allowed the user to monitor around 200-250 sensors at once, with the usual detail displays that provide specific localized information. Each sensor monitored a node in the processing network, although the main display, using a group alarm approach (a type of statistical summary), provided very little information other than the most serious condition found in the underlying detailed data. However, no information was provided about what was causing errors, or even where in the subset errors were occurring. The users were having a difficult time understanding how the network was behaving.

Despite the paucity of information provided by the group alarm technique, the main display was nearing its upper limit on the number of sensor monitors that could be placed on the screen at one time. Still, users foresaw the need to monitor many more sensors at one time (into the thousands), and realized that all forms of scaling the current display design would be insufficient. The users were already struggling to understand the network, and adding more sensors to monitor would have exacerbated the problem.

The large information set demanded that the display allow users to refocus their attention to important events in the network, which is best achieved in a longshot display. The new display allows operators to monitor tens of thousands of data points at one time, in real time, by taking advantage of Micro/Macro design principles [4].

The data reported on each system is placed into a single 2 pixel high row. There are four main components to the network: the inboxes that receive data, the software that processes data, the outboxes that send data, and the overall hard drive of the system. In general, inboxes or outboxes that are full, overworked/down processes, and full hard drives are problematic. Data that are normal are less important (although must be shown). Within the display, as inboxes and outboxes grow fuller, the data mark representing their value moves towards the outside of each row (breaking away from the rest of the data) and become brighter on the screen. Similarly, as processes and hard drives moved from nominal, they became more prevalent on the screen (changing color and becoming brighter). The user is quick to see where problems are occurring, since they almost leap off of the display, even though each mark is only 2 pixels wide by 2 pixels high. When multiple problems are occurring at the same time, the user is able to know where to focus, since the type and magnitude of each problem is known. Once a user decides to focus on a particular network issue, Performance View allows the user to quickly access a detailed display for the system of interest.

the visualization. In doing so, it is clearly visible which reports are coming due and which are not, which will help the analyst decide which report to work on next.

However, the reader is cautioned from thinking that the same frames of reference easily and arbitrarily generalize across many visualizations; the most important information required to complete the cognitive work should always determine the appropriate frames of reference [2]. Returning to the preceeding example, the importance of the analyst report (i.e., priority) and number of previous reports generated on that topic may be more important than the report's due date, and the representation designer may choose to use these as the frames of reference. There is no generic way to generate the proper frames of reference—the designer must understand the cognitive work of the analyst and determine the relative importance of all the information required to support that work. Techniques exist to support the extraction and distillation of this cognitive work. Within the field of Cognitive Systems Engineering, the application of a Cognitive Work Analysis has traditionally been the technique by which our understanding of work domain demands is turned into representation design requirements [3].

Establishing the proper frames of reference for the data is important to place the data in its proper context, but a longshot display also must allow an analyst to focus and refocus on the data of interest. The understanding of the data set as a whole must be offset with the ability to understand an individual datum within the set. This presents the user with two perspectives on the data that they must handle, labeled by Tufte as the Micro/Macro approach [4], in which the representation of each datum allows the data to group into a larger whole.

The concept of combining detailed, focused central processing within the context of a broader surround is so fundamental it forms the basis of our own visual processing in the eye. Because the world presents far more data than we can analyze in detail, our visual system is organized into a central region with very high acuity and a surround region that is processed at lower resolution. Our visual field is ap-

proximately 180 deg, but only the central 2 deg is processed in the highest level of detail. This center-surround arrangement enables humans to be aware of significant events (e.g., a fast moving object) over a very large field of view, and then move our eyes to focus the high acuity processing on the region of interest. If our eyes only possessed the relatively small 2 deg region of detailed viewing, we would never know when potentially important events had happened outside this area of focus. Similarly, visualization systems that contain only detail views with no corresponding longshot display do not sufficiently support analyst refocusing.

This ability for an analyst to understand both the whole and the individual is a key benefit provided by longshot displays. While analysts can only work on a single datum (or a few data) at a time, the dynamic nature of world events and analytic tasking means that the appropriate focus may be constantly changing. Important new events are constantly occurring (or at least may be occurring), and visualizations must support the analyst's need to refocus even as they are engaged in detailed consideration of a narrow part of the data set. In showing the entire data set, longshot displays enable analysts to refocus when events change in the world, to hone in on the specific data that should be worked on given the current situation (see: Longshot Example 2). When events change, the analyst is easily able to see what aspect of the data set is the most important and decide whether or not to redirect the focus based on the changing situation. This ability to redirect attention is vital in dynamic, time-critical situations.

Supporting an analyst's ability to focus and refocus within the data set helps overcome what has been labeled the Keyhole Effect, which occurs when people are only provided a limited view from the much larger information space of relevant information². The Keyhole Effect hobbles a viewer's peripheral awareness by narrowing both perceptual and attentional mechanisms, leading analysts to miss new



Figure 1: Without a longshot display, systems force users to view their data through a keyhole. Longshot displays provide one way of escaping the Keyhole Effect by placing data in the necessary context so that users can determine where in the data set to focus (photo courtesy of Flip Phillips).

events, have increased difficulty in navigating novel data environments, and generate incoherent models of the explored data space [5].

This Keyhole Effect is instantiated in many different ways, the most common—though not only—being the scrolled data list (or any table, spreadsheet, etc. that includes a scroll bar to see the entire set). Displays that show only a focused, key-holed view of data give analysts great detail into a select few data while giving the analyst no insights into the remainder of the data set. This display makes it virtually impossible for analysts to redirect their attention when an important event happens outside their focused view. In these visualizations, the representation designer is requiring the analyst to look at the right place at the right time to see when changes occur. The larger the data set, the more unlikely that the analyst will ever be fortunate enough to have their detail view pointed at exactly the right place to see the onset of a significant event.

However, longshot displays, with all the data presented simultaneously, make it easier for analysts to see where they

should be focusing their efforts. Analysts can more effectively manage their time and effort working with the data, ensuring the most important work is getting done at all times.

This ability to focus and refocus is a key difference between longshot displays and displays that allow analysts to completely zoom out on a data set (designers often build this capability into maps). While all the data are visible on the screen, the zoomed out view simply jams the data into the window, making most of the data incomprehensible. Such displays leave analysts unable to determine where to focus, requiring them to zoom in and scan around to determine the appropriate region of focus, thus defeating the purpose of the zoomed-out view.

In the same vein, statistical summaries of a data set do not sufficiently allow a user to refocus in the data set. While each data value is taken into account in a statistical summary, the actual data values are hidden from the user—producing a different type of keyhole. While the statistical summary can allow judgments on the data set as a whole, it prevents data judgments at the individual level. Each datum has a value unto itself, and this value needs to be known, especially since analysts do not work on the data set as a whole—they work on an individual datum. It becomes extremely difficult to work on a single datum when it is rolled up into a summary, and it certainly will not allow the analyst to adjust focus when necessary.

Longshot displays, on the other hand, make it possible to work with each individual datum as a true, independent entity. Additionally, longshot displays make it possible to understand the impact of outliers on the data set. As a small subset of the data set, outliers are not able to greatly impact the status of the overall data set. Yet, these outliers are often the interesting data to analysts, as they represent where important events are occurring. With longshot displays, analysts are able to focus on both the typical data and the atypical data, depending on their current interests, mak-

² The Keyhole Effect refers to the ability of a person to understand a room based on what they can see by looking through a keyhole in a door. A person can focus on a small part of the room, and can even scan around the room, but can never truly understand what is going on in the room through this keyhole view. In the same way, analysts will not be able to understand their entire data set while they are narrowly focused on only one small portion of it.

ing selection of data to attend to easier.

This need to refocus also has implications for the use of filters on the data set. When large data sets are collected, data filters are used to reduce the data set available to analysts by eliminating data that are deemed irrelevant. Filters can be a very useful method for supporting analysis of extremely large data sets, but they can also be misused and ultimately create additional work for analysts.

System designers use filters based on the premise that it is possible to know a priori what data will be important and what data will be irrelevant. When this distinction can be made, it becomes easy to separate what seems to be good data from what seems to be bad data. Unfortunately, filters are often designed to completely eliminate those data that are deemed irrelevant. This is the ultimate form of a keyhole. Not only are analysts given only a partial view of the data set at a time (since once data are filtered, they are often placed into standard representation keyholes), but there are large portions of the data set that are left completely irretrievable. This complete removal of data causes a couple of complications for analysis. First, analysts have no idea what data have been removed by the filter. While the distinction between important and irrelevant data may be clear some of the time, just as often the difference between important and irrelevant is very murky. This is especially true in intelligence collection, in which obfuscation can be a key tactic of an adversary. Unfortunately, since the filter removes the data, analysts will have no way of knowing what they are not seeing—the data filter prevents users from ever refocusing on these data. These data can never be examined more closely. While the majority of the data may indeed be irrelevant, the select few that are relevant may be most valuable to the analysts.

Since longshot displays are able to represent large data sets to the analyst, there is less of a need to completely remove data that does not pass the filters limits³. Instead of removing the data, the

representation designer can highlight the most relevant data and reduce the visibility of the less relevant data, making it easier for the analyst to focus on what is important while still preserving the ability to see the larger context. Representation designers can control the color, brightness, size, and position of each datum (to name a few options) in order to give analysts insights into how important each datum may be. Regardless of how important each datum seems, all data make it to the screen, and analysts can still return to those data if they want to broaden their focus. This gives analysts greater control over their data set and can allow them to see disconnects between what a pre-programmed relevancy generator has deemed to be irrelevant for the task and what the analyst actually thinks is irrelevant.

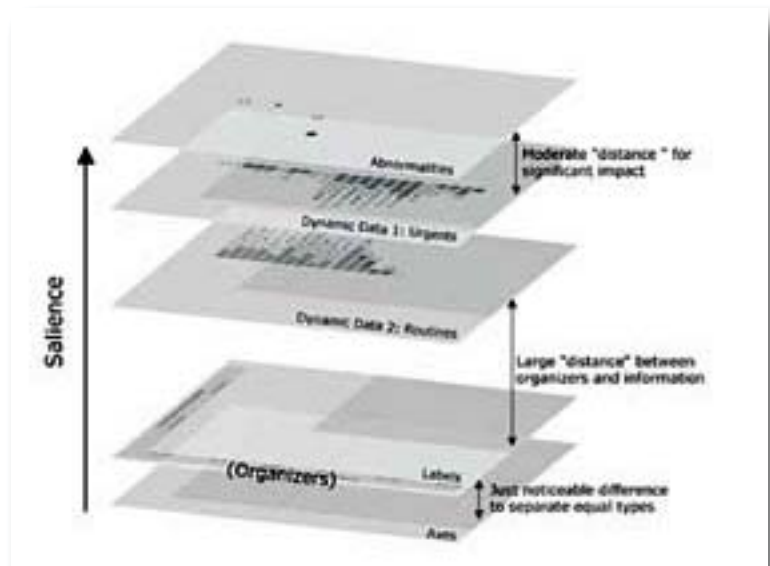
A useful technique for achieving some of the information highlighting described above involves utilizing the natural perceptual way in which our visual system processes different types of visible features to influence the perceived salience of particular graphical elements of the display. A salience map (see Figure 2) enables a representation designer to make a specific rank ordering of the various elements that will constitute the final display. The overall display space can be structured based on decisions about the relative importance of information related to the cognitive work the display must support.

The appropriate number of levels in a salience map depends on the complexity of the domain under consideration and the nature of the graphical elements in the display design. In the example salience map shown in Figure 2, we show five salience levels. At the lowest level

of salience are graphical elements that organize and provide structure for the overall display (e.g., Plot labels). At the next level are so-called static data or reference marks. These are important values for the system, but they are known in advance and do not change in the course of normal operations. The last three levels of salience are reserved for dynamic data elements: (1) routine dynamic data should be the least salient of the three levels, with more salience assigned to (2) urgent dynamic data, and finally, (3) abnormalities or alarms should be the most salient items on the display.

Salience in the display is a statement of design intent which results from an assessment of the relative importance of the different information required to support the relevant cognitive work (e.g., abnormalities or unexpected events should be made most salient on the display in order to immediately draw the attention of the operator). However, perceived salience of the visual display is based on the pre-attentive perceptual mechanisms of the human visual system, and results from spatial and/or temporal contrasts of the individual display elements [6]. Thus, in order to effectively direct the analyst's attention, create relevant contrasts to highlight information, and group data into meaningful combinations, representation designers must

Figure 2: The use of a Salience Map to structure information in a visual display. Controlling salience allows a representation designer to effectively direct attention and highlight information based on the cognitive work the display must support.



³ In some cases, filters are unavoidable. Even when some data must be removed from the data set altogether, this does not free the designer from the responsibility of using a longshot display for the remaining data set.

understand and utilize characteristics of our pre-attentive perceptual mechanisms.

Finally, the ability to focus and refocus within the data set provides little value if the analyst is unable to work on a particular datum. Therefore, in order to satisfy the final criteria, the longshot display must also be coupled with a detail display that allows users to work on the focused data. Longshot displays are important, but they cannot exist in isolation, as they will support some of the user's cognitive work, but cannot support all of it. The two data representations (longshot and detail displays) provide complementary perspectives on the same data and will be of great benefit when combined into a single tool for analysts. The longshot display provides analysts with the ability to see the entire information space to determine where to focus, while the detail display allows the analyst to focus in on the area selected. Combining the longshot and detail views helps analysts complete the difficult cognitive work they face, and allows analysts to quickly go from focusing on the data set as a whole to focusing on a single datum to assess and then back again as needed.

Supporting a New Analytic Approach for Working through Large Data Sets

Longshot displays are critically necessary complements to existing detailed views in order to add the robustness of directing/redirecting attention to the already powerful 'drill down and inspect' capabilities provided by traditional detail displays. They support additional cognitive work that cannot be supported solely by detail displays (e.g., refocusing when events change and selecting the proper data for analysis given the situation), allowing analysts to alter their approach to large data sets. Rather than using the FSA model, analysts can shift to the richer and more resilient Organize-(Re)Focus-Analyze (OFA) model of intelligence analysis.

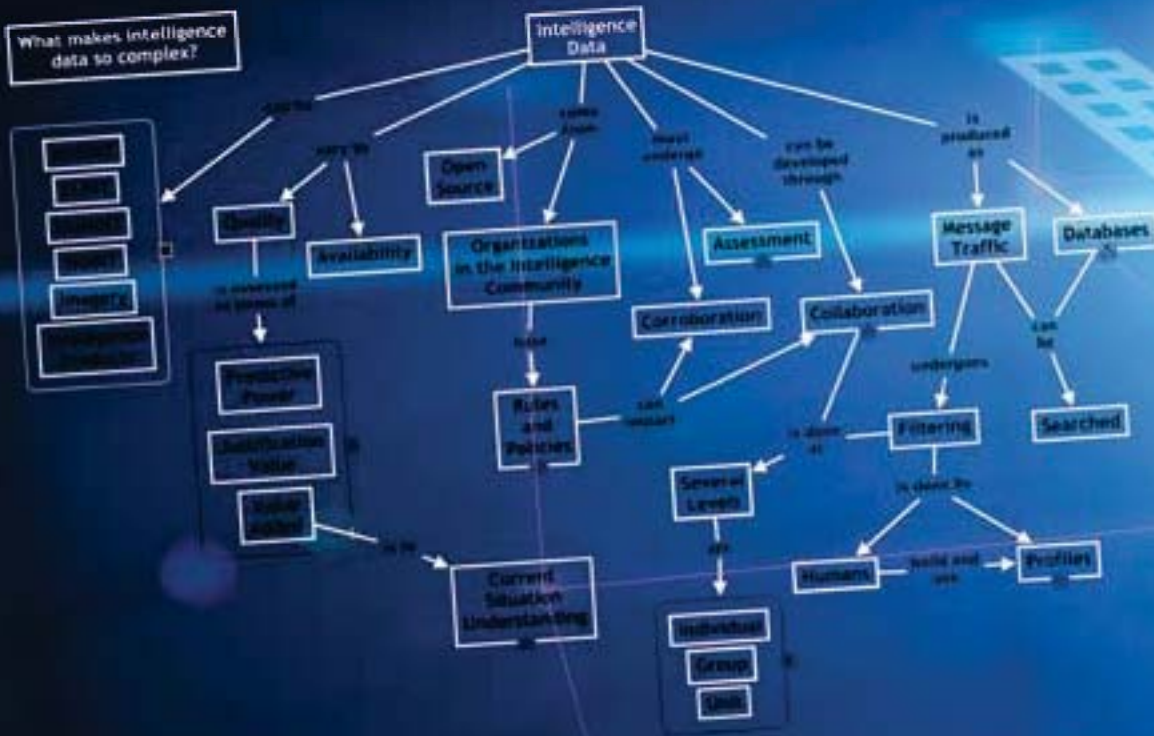
By choosing the proper frames of reference for the data, the longshot display utilizes the human perceptual system's tendency to detect patterns in the world, organizing the data in a meaningful con-

text, and allowing the analysts to focus on the portion of the data as desired. As the pattern changes, so too can the analyst's focus. This can enable better navigation of the information space, allowing analysts to work on the most important data. With tightly coupled longshot and detail displays, analysts are able to understand and manage the entire data set, and still generate an analysis on the individual datum. By adding a longshot display to the detail display, analysts' work will never be limited by what may be just outside their focused view—everything is in view. ■

References and Selected Reading

- [1] Patterson ES, Roth EM, & Woods DD. Predicting Vulnerabilities in Computer-Supported Inferential Analysis Under Data Overload. *Cognition, Technology & Work*. 2001;(3):224-237.
- [2] Norman DA. *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Reading (MA): Addison-Wesley; 1993.
- [3] Elm WC, Potter SS, Gualtieri JW, Roth EM, & Easter JR. Applied Cognitive Work Analysis: A Pragmatic Methodology for Designing Revolutionary Cognitive Affordances. In: Hollnagel E, editor. *Handbook for Cognitive Task Design*. London(UK): Lawrence Erlbaum Associates, Inc.; 2003.
- [4] Tufte ER. *Envisioning Information*. Cheshire (CT): Graphics Press; 1990.
- [5] Woods DD, Watts JC. How not to have to navigate through too many displays. In: Helander MG, Landauer TK, Prabhu P, editors. *Handbook of Human-Computer Interaction*. 2nd ed. Amsterdam (NL): Elsevier Science; 1997. p. 617–650.
- [6] Wolfe JM. Visual Search. In: Pashler H, editor. *Attention*. London (UK): University College Press; 1998.

Knowledge Visualization with Concept Mapping Tool



Mission-critical knowledge resides in the memories of experts and in a relatively small subset of external resources, often widely distributed, such as technical reports, manuals, books, online material, etc. In order to make such knowledge explicit and sharable, it is necessary to have an effective elicitation and capture methodology with a powerful and useful representation scheme.

There are many methods and tools that can perform one or more of these requirements. Through research, experimentation, and observation, achieving many of the requirements led to the discovery of a simple yet powerful representation scheme (concept mapping) based on Ausubel's Assimilation Theory of meaningful learning [1]. The representation scheme described by Joe Novak is based on Ausubel's Theory, which is different from other methodology schemes such as mind maps, knowledge maps, conceptual structures [2], and the like, in that it conveys rich semantic content and also provides an organizing factor for other resources and media that are added to the concept map knowledge model.

Background

Concept maps were developed in 1972 in the course of Novak's research program at Cornell, where he sought to follow and understand changes in children's knowledge of science [3]. During the course of this study, the researchers interviewed many children and found it difficult to identify specific changes in the children's understanding of science concepts by examination of interview transcripts. This program was based on the learning psychology of David Ausubel [1][4][5]. The fundamental idea in Ausubel's cognitive psychology is that learning takes place by the *assimilation* of new concepts and propositions into existing concept and propositional frameworks held by the learner. This knowledge structure as held by a learner is also referred to as the individual's *cognitive structure*. Out of the necessity to find a better way to represent children's conceptual understanding emerged the idea of representing children's knowledge in the form of a concept map. Thus was born a new tool not only for use in research, but also for many other uses.

The Foundation of Concept Map

Living in the digital age made it necessary to transform the meaningful learning methodology into a more desirable format. Therefore, the Institute for Human and Machine Cognition (IHMC) with the help of Joe Novak created a computer-based tool that included all foundational theories. The Cmap Tool was designed to be a simple-to-use but powerful environment that facilitates collaboration and sharing during the construction of a concept map-based knowledge model. Concept maps are meaningful diagrams (directed graphs) for organizing and representing knowledge by using *concepts* and *propositions* as node-link-node expressions of the simple form [6].

<A, r, B>

A concept is defined as a perceived regularity in events or objects, or records of events or objects, designated by a label. The label for most concepts is a word, al-

though sometimes multiple words or symbols such as + or % are used. Propositions are statements about some object or event in the universe, either naturally occurring or constructed [7]. Propositions contain two or more concepts connected using linking words or phrases to form a meaningful statement. Sometimes these are called semantic units, or units of meaning.

Characteristics of Concept Maps

Concept maps are depicted by the concepts that are represented in a hierarchical fashion with the most inclusive, most general concepts at the top of the map and the more specific, less general concepts arranged hierarchically below. The hierarchical structure for a particular domain of knowledge also depends on the context in which that knowledge is being applied or considered. Therefore, it is best to construct concept maps with reference to some particular question, called a *focus question*, for which an answer is sought. A situation or event that requires organization of knowledge for understanding may provide context for the concept map [8].

Another important characteristic of concept maps is the inclusion of *cross-links*. These are relationships or links between concepts in different segments or

domains of the concept map. Cross-links help one see how a concept in one domain of knowledge represented on the map is related to a concept in another domain shown on the map. In the creation of new knowledge, cross-links often represent creative leaps on the part of the knowledge producer.

There are many more innovative characteristics to concept maps, for example, in the case of *resources*, which are events or objects that will help clarify the meaning of a given concept. Normally these are not included in ovals or boxes, since they are specific events or objects and do not represent concepts. Resources can be anything, for example, images, videos, URLs, spreadsheets, PowerPoint slides, documents, other concept maps, etc. It is critical to understand that these and many more characteristics of concept maps are important also to the facilitation of creative thinking and innovation.

Applications of Concept Maps

To give a sense of the utility of Cmaps, they have been used to show gaps in trainees' knowledge, capturing the proficiency scale for a level of performance. Concept mapping is used as part of the curricular infrastructure of many individual schools

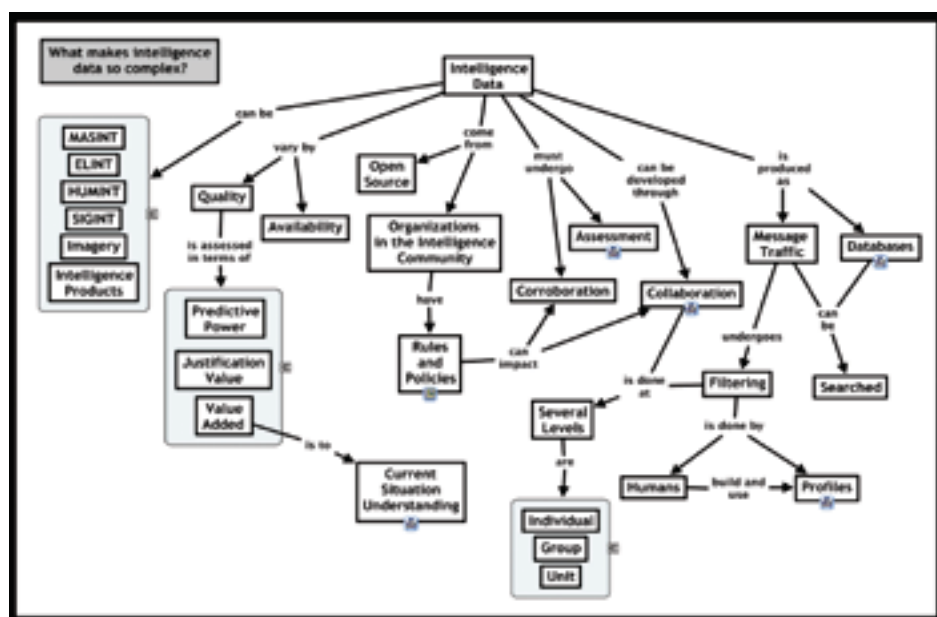


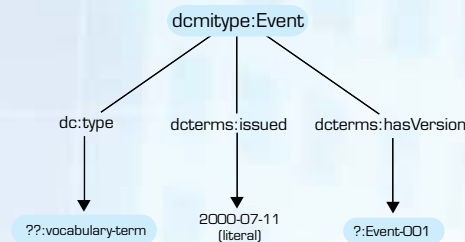
Figure 1: An example of a concept map. This map was made on the basis of collective external experience after conducting a cognitive task analysis with intelligence analysts, in an unclassified setting.

Concept maps can be used to:

- elicit knowledge from experts
- design new technology by domain experts
- reveal expert-novice differences
- acquire software-assisted knowledge
- brainstorm
- share knowledge
- support training and performance
- identify knowledge gaps
- create new knowledge, for example, turning tacit knowledge into an organizational resource
- map team knowledge
- structure conceptual queries
- compare methodologies
- structure linguistic definition
- design competency questions

and by entire school systems [9]. Concept map knowledge models can serve as living repositories of expert knowledge, encouraging knowledge sharing as well as supporting knowledge preservation [3]. They can serve as interfaces for decision support systems where the model of the expert's knowledge becomes the interface for a performance support tool or training aid [6]. Concept map knowledge models have successfully captured and preserved the domain knowledge of experts in such areas as terrain analysis, aviation weather forecasting, and engineering [9].

Through first-hand knowledge and observation, there were many more interesting applications of concept maps. As more and more users find their niche, there have been great benefit and promise towards mission in the usage of the concept mapping tool and its fundamental methodology. It brings together communities to share and collaborate on knowledge.



```
<owl:Class rdf:ID="Choice">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#ControlConstruct"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:ObjectProperty rdf:about="#components"/>
      </owl:onProperty>
      <owl:allValuesFrom>
        <owl:Class rdf:about="#ProcessComponentBag"/>
      </owl:allValuesFrom>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

Figure 2: A representation of what a very simple RDF triple would look like visually, and, on the right, is a textual representation in OWL.

The wide use of concept mapping from grade school kids to research scientists has created many architectural challenges for the Cmap Tool. The present architecture has been relatively closed to third-party developers who wish to interoperate with the Cmap Tool until recently, when a new approach called the Knowledge Exchange Architecture (KEA) [10][11], provided the data and procedural framework to allow third-party developers to design programs that interact with the Cmap Tool suite of programs, thus providing a new range of applications that can be developed and integrated with Cmaps.

Formalizing the Informal of Concept Maps

Upon first seeing concept maps, a common reaction is, "That's just a flow diagram." Only with time, instruction, and experience do people come to see that concept maps can take on different functions with respect to the representation of concepts and their relations. Upon further observation and research, the development of expressing concept maps for the representation of artifacts used in the Semantic Web (Web 3.0) were created, more specifically the Web Ontology Language (OWL) [12]. The confirmation that much of the utility of the Cmap Graphic User Interface (GUI) arises from its ability to utilize the essentially spatial structure of the display in meaningful ways, not the least simply as a memory aid and a rapid visual recognition mode for "familiar" knowledge

structure. The spatial (diagrammatic) and temporal features of Cmaps (Cmapping as a process) can be used as a tool to achieve clarity about intended meanings of propositions (concepts and their relations).

The adoption of the Cmap Tool interface to a Semantic Web Graphic User Interface (GUI) provides a fully functional display, editing and recomposing of OWL, and Resource Description Framework (RDF) meta models. The tool, called CMap-based Ontology Editor (COE), is acquiring widespread use as an ontology viewer and editor, capable of importing and exporting various formats. COE enables users to display, manipulate, edit, and compose ontologies and other formal representations in a variety of languages.

The challenge to display very large ontologies provided insight into the overall look and feel of understanding concept maps. The idea of visualizing, the ability to handle rapid navigation facilities and "zooming" provided some novel approaches in this domain. This technique has enhanced the overall capability of the Cmap Tool. The Semantic Web community has taken great interest in the visualization of these formal languages to gain insight from the normal textual method of descriptions [13].

COE has been used to interface with knowledge mining engines and reasoners. It has been adapted to input useful data and to display results from data analysis. The key to moving from informal to formal

representation allows one to see the possibility of a smooth transition of automation without needing to learn the details and to use the machine to help offload many functions as well as share the information.

Building this bridge between knowledge models that reflect human learning and formal ontologies that can support machine inference is a challenge, but it is necessary and important.

Summary and Conclusion

The theoretical foundation and the origins of concept maps have been presented in this article. Initially, concept maps may appear to be just another graphical representation of information. Understanding the foundations of concept maps and proper use of the Cmap Tool will lead the user to see that it is truly a profound and powerful tool. The software was designed to have a low threshold, making it easy to learn for new and naïve users, but at the same time have a high ceiling, enabling sophisticated and expert users to develop large, complex knowledge models. Concept mapping's greatest contributions are facilitating collaboration, documenting team analytic efforts, targeting information, and fostering an analytical environment.

References

- [1] Ausubel DP. *The Psychology of Meaningful Verbal Learning*. New York (NY): Grune and Stratton; 1963.
- [2] Sowa JF. *Conceptual Graphs as a Universal Knowledge Representation*. *Computers & Mathematics with Applications*. 1992;23(2-5):75-93.
- [3] Novak JD & Gowin DB. *Learning How to Learn*. Ithaca (NY): Cornell University Press; 1984.
- [4] Ausubel DP. *Educational Psychology: A Cognitive View*. New York (NY): Holt, Rinehardt and Winston; 1968.
- [5] Ausubel DP, Novak JD, & Hanesian H. *Educational Psychology: A Cognitive View*. 2nd ed. New York (NY): Holt Rinehart and Winston; 1978.
- [6] Hoffman RR & Linton G. *Eliciting and Representing the Knowledge of Experts*. In: Ericsson KA, Charness N, Feltovich P, & Hoffman R, editors. *Cambridge Handbook of Expertise and Expert Performance*. New York (NY): Cambridge University Press; 2006. p. 203-222.
- [7] Novak JD & Cañas AJ. *The Theory Underlying Concept Maps and How to Construct and Use Them*. Pensacola (FL): Institute for Human and Machine Cognition; 2008. Report No.: IHMC CmapTools 2006-01 Rev01-2008
- [8] Cañas AJ & Novak JD. *Re-examining the Foundations for Effective Use of Concept Maps*. In: AJ Cañas & JD Novak, editors. *Proceedings of the Second International Conference on Concept Mapping*; 2006. San Jose, Costa Rica: University of Costa Rica. p. 494-502.
- [9] Cañas AJ, Coffey JW, Carnot MJ, Feltovich P, Hoffman R, Feltovich J, & Novak JD. (Institute for Human and Machine Cognition, Pensacola, FL) *A Summary of Literature Pertaining to the Use of Concept Mapping Techniques and Technologies for Education and Performance Support*. Report to the Chief of Naval Education and Training; 2003.
- [10] Cañas AJ, Hill G, Granados A, Perez C, & Perez JD. *The Network Architecture of CmapTools*. Pensacola (FL): Institute for Human and Machine Cognition; 2003. Report No.: IHMC CmapTools 2003-01.
- [11] Cañas AJ, Hill G, Bunch L, Carff R, Eskridge T, & Perez C. *KEA: A Knowledge Exchange Architecture Based on Web-services, Concept Maps, and CmapTools*. In: Cañas AJ., & Novak JD, editors. *Proceedings of the Second International Conference on Concept Mapping*; 2006. San Jose, Costa Rica: University of Costa Rica. p. 304-310.
- [12] Eskridge T, Hayes P, Hoffman R, & Warren, M. *Formalizing the Informal: A Confluence of Concept Mapping and the Semantic Web*. In: Cañas A, editor. *Proceedings of the Second International Conference on Concept Mapping*; 2006. San Jose, Costa Rica: University of Costa Rica.
- [13] Verirl I. *What is the Value of Graphical Displays in Learning?* *Education Psychology Review*. 2002;(14):261-290.

Visualizing Data in the Metaverse

From the earliest cave drawings of a mastodon hunt to Leonardo Di Vinci's sketches of the ornithopter, humans have tried to visualize their experiences and their dreams. Images captured on cave walls provided a collective perspective of the hunt that might have altered tactics for the next kill; detailed renderings of feathers and fulcrums helped turn flights of fancy into flying machines. People often need to *see* how things work before we *understand* how they work.

Throughout history, new ways to visualize data have repeatedly spawned discovery and invention. Through the twentieth century, these visions were captured in drawings, diagrams, photographs, and even movies. But data visualization has evolved into more than a “snapshot” of an event or physical state to preserve it for analysis. New data visualization tools can provide a portal for the viewer to step through, to enter and experience an otherwise inaccessible world.

As the ways we visualize information rapidly evolve, so do the ways we access that information. It took only one generation for libraries of atlases and textbooks to yield to shared databases and online access. Now, pages are turned by a clicked mouse button instead of a licked thumb. It may take even less time for scrolling through pages on a computer monitor to become equally arcane, as the computer world we know today yields to the virtual world of tomorrow.

A digital virtual world—the *Metaverse*—was vividly described by Neal Stephenson in his 1992 cyberpunk novel *Snow Crash*. Residents of the Metaverse buy, sell, socialize, create, and learn, just as they do in the physical universe. The only real difference is that these residents are digital manifestations of the people who create them—control them—are them.

Online games such as *World of Warcraft* and *Everquest* have prepared a

generation of gamers for the virtual world that may be redefining how we work, learn, shop, and play. Just as data models and simulations have provided new ways to visualize substances and systems, the metaverse gives us an alternative way to look at ourselves and the environment we inhabit.

Second Life was the first virtual world to demonstrate what life in the metaverse could be like. Launched in 2003 by Linden Research, Inc., Second Life (SL) slowly amassed a devout following, registering over 13 million accounts by March 2008. At any given time, tens of thousands of SL residents can be found engaged in virtually every activity you would expect to find among a similar population in real life.

The curators of SL, at Linden Labs, are committed to letting their virtual world evolve naturally, with communities, customs, laws, and even an economy settling into social equilibrium with few constraints. Their new world was designed to be created by its residents—a crucible for a grand social experiment.

The first to come to SL were the gamers. They were soon followed by profiteers, hucksters, and people who were just curious. As SL has grown in popularity—and profitability—businesses, merchants, government agencies, nonprofit organizations, museums, and schools have started setting up shop as well. Along with the virtual beach houses and nightclubs,

the metaverse is filling up with more serious projects that lend themselves to interactivity.

Businesses have opened offices in SL to conduct virtual meetings and sales presentations. Retail merchants are using SL to promote their products, with some shops set up to let customers design virtual products for their avatars, and then receive similar items for themselves. Government agencies, embassies, and various service industries have discovered the benefits of providing easy, interactive access to their information, any time of day or night. Several museums have found that 3-D exhibits in SL are a natural extension for their local displays. And the potential for education and research is boundless, with numerous schools and universities already building virtual campuses and hosting classes in SL.

As supercomputing power becomes almost commonplace, elaborate data models and simulations are finding their way into more and more hands. The increase in accuracy and resolution of what we perceive provides for better interpretations of how things work and what things mean. These new ways to present information are being met with new ways to receive it. Visualizing data does not have to be limited to a passive event in two-dimensional space. In the metaverse, the possibilities for visualizing data are virtually limitless. 📺

